The Examiner further states that S.N. 08/202,236 and S.N. 08/177,920 (parent application of the instant application S.N. 08/475,822) are divisional applications from the parent application S.N. 07/158,652. However, the Examiner states that "there is no evidence that the PTO has set forth a restriction requirement between the nucleic acids of application Serial No. 08/202,239 and the methods of use of those nucleic acids as probes in the parent application Serial No. 08/177,920.

Since this rejection is provisional, Applicants respectfully request that the Examiner hold in abeyance the instant rejection. Upon the indication of allowable subject matter in the instant application, Applicants reserve the right to file a terminal disclaimer or traverse the rejection.

The specification is objected to and claims 11-18 are rejected under 35 U.S.C. § 112, first paragraph, as the specification allegedly fails to adequately teach how to make and/or use the invention, i.e., fails to provide an enabling disclosure.

The Examiner states that the specification does not teach how to use the invention for the claimed diagnostic methods, which include the nucleic acids of ORF-Q, ORF-R, ORF-1, ORF-2, ORF-3, ORF-4, and ORF-5, as claimed herein. Allegedly, the nucleic acid hybridization with HIV-1 to assay HIV-1 is allegedly not demonstrated. More particularly, the Examiner states that the conditions and methods are not given to distinguish HIV-1 from other retroviruses. The Examiner cites Hahn et al. as demonstrating that at the time of filing the instant invention, it was known that cross-hybridization occurs between the sequences of HIV and members of the HTLV family. The Examiner concludes that in view of the specification's alleged lack of sufficient teachings of specific hybridization using the claimed probes and Hahn et al., which shows cross-

LAW OFFICES
FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER, L.L.P.
1300 I STREET, N. W.
WASHINGTON, DC 20005
202-408-4000

hybridization to members of the HTLV family, the specification is non-enabling for the claims. Applicants respectfully traverse the rejection.

The PTO has the burden of establishing a *prima facie* case of lack of enablement. Furthermore, applicants' specification disclosing how to make and use the claimed invention must be taken as in compliance with § 112, first paragraph, unless there is a reason to doubt the objective truth of the disclosure. In re Brana, 51 F.3d 1560, 1566, 34 U.S.P.Q.2d 1437, 1442 (Fed. Cir. 1995); *citing* In re Marzocchi, 169 U.S.P.Q. 367, 369 (C.C.P.A. 1971).

Applicants respectfully submit that the specification provides the necessary guidance to teach one of skill in the art how to use the claimed invention. More particularly, it is clear from the specification and the skill in the art that one would appreciate that the nucleic acid probes of ORF-Q, ORF-R, ORF-1, ORF-2, ORF-3, ORF-4, and ORF-5, are indeed capable of detecting the presence or absence of HIV-1.

Based on such teachings, applicants submit that the enablement requirement is met. Indeed, the 35 U.S.C. § 112, First Paragraph, Enablement Training Manual, August, 1996, provides that:

> Unless a specification specifically states something to the contrary, the term "diagnostic assay" is to be construed to mean any assay that can be used to help diagnose a condition, as opposed to an assay that can, in and of itself, diagnose a condition. . . Therefore, to enable a diagnostic assay use, a disclosure merely needs to teach how to make and use the assay for screening purposes.

(Id at 22-23.) Here, the specification provides that "all of the above mentioned peptides can be used in diagnostics as sources of immunogens or antigens free of viral particles." (Specification at 16, lines 6-8.) The hybridization assays for the detection of HIV-1 are set forth at page 14, line

LAW OFFICES
FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER, L.L.P.
1300 I STREET, N.W.
WASHINGTON, DC 20005
202-408-4000

11, through page 15, line 8. Therein, applicants teach that hybridization techniques were well-known in the art at the time the application was filed. It is stated that "[u]sing the cloned DNA fragments as a molecular hybridization probe - either by marking with radionucleotides or with fluorescent reagents - LAV virion RNA may be detected directly in the blood, body fluids and blood products [] and vaccines . . . " (Specification, page 14, lines 17-22.)(Parenthetical removed.) For example, applicants teach that hybridization assays using nucleic acid probes for Hepatitis B virus were known in the art. (Specification at page 14, lines 29-32.)

Further support for the knowledge in the art at the time the claimed invention was made is found in Arya et al., "Homology of Genome of AIDS-Associated Virus With Genomes of Human T-Cell Leukemia Viruses," Science, 225:927-930 (August 31, 1984) (Exhibit 1). Therein, the authors exemplify hybridization experiments between HTLV-I and -II, and HTLV-III.

In addition, Hahn et al., cited by the Examiner, further depict the use of an HTLV probes in hybridization assays. (Hahn et al. at 168.) It is noted that the Examiner relies upon Hahn et al. to teach the cross-hybridization between sequences of HIV and HTLV, even in stringent conditions. (Paper No. 20, at 4.) However, Hahn et al. discuss that the complete genomes of HTLV-I, HTLV-Ib, and HTLV-II were digested with restriction enzymes and hybridized with the full-length of HTLV-III probe in "relaxed conditions." (Hahn et al., page 168, second column, lines 3-9.) In particular, the legend of Figure 4 indicates that low stringency hybridization of 8 X SSC, 20% formamide, 10% dextran sulphate at 37 °C, and washing conditions of 1 X SSC from 22-65 °C were used. The fact that Hahn et al. use a low stringency hybridization indicates a desire to cross-hybridize with other sequences. Low stringency, one of ordinary skill in the art

LAW OFFICES
FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER, L.L.P.
1300 I STREET, N. W.
WASHINGTON, DC 20005
202-408-4000

would know, generally leads to greater cross-hybridization results. By using the low stringency hybridization conditions, Hahn is obviously attempting to show cross-hybridization between each of the "members of the HTLV family." The text of Hahn, at the bottom of page 168, indicates such an attempt in order to "evaluate sequence homology."

Thus, Hahn does not attempt and cannot be read to show whether or not specific hybridization is possible with the claimed invention herein. Hahn et al. simply did not attempt an experiment, which could show that possibility. Figure 4 of Hahn, therefore, cannot support the Examiner's conclusion.

On the other hand, Alizon et al., in "Molecular Cloning of Lymphadenopathy-Associated Virus," Nature 312:757-760 (1984) (Exhibit 2), describe the discriminating hybridization assays using a probe specific for HIV-1. Therein, high stringency hybridization conditions of 50% formamide and 5 x SSC at 42 °C and washing conditions of 0.1 X SSC at 68 °C were used. Therefore, it would have been readily appreciated that the determination of the hybridization conditions is well within the purview of the skilled artisan and dependent upon the goal of the particular research. No reasonable evidence to suggest that the nucleic acids recited in the claims could not discriminate between different retroviral DNA sequences has been presented by the Examiner.

To the contrary, applicants submit that the claim-designated nucleic acids are unique to HIV-1. Therefore, one having skill in the art would acknowledge the use of such nucleic acids as probes in hybridization assays.

LAW OFFICES
FINNEGAN, HENDERSON,
FARABOW, GARRETT
& DUNNER, L.L.P.
1300 I STREET, N.W.
WASHINGTON, DC 20005
202-408-4000

For example, ORF-Q corresponds to *vif* protein of HIV-1. The *vif* protein (virion infectivity factor) is also known as *sor*, *A*, *P'*, and *Q*. (Gallo et al., "HIV/HTLV Gene Nomenclature," Nature 333:504 (1988) (Exhibit 3).) The *vif* protein is not found in HTLV-1 or HTLV-II (Gallo et al. at 504), and therefore, a nucleic acid probe corresponding to this protein would not detect these viruses in a hybridization assay.

Furthermore, although a *vif* protein is present in the genome of HIV-2, the nucleotide sequences of the *vif* proteins of HIV-1 and -2 have only about 45% homology. This is shown by a comparison of the nucleotide sequence of ORF-Q of HIV-1 given in applicants' specification with the nucleotide sequence of ORF-Q of HIV-2 (i.e., *vif*) given in Guyader et al., "Genome Organization and Transactivation of the Human Immunodeficiency Virus Type 2," Nature, 326:662-669 (1987) (Exhibit 4). Exhibit 5 shows the nucleotide sequence comparison of the two sequences. Because there is only about 45% homology between the nucleotide sequences of the two proteins, a nucleic acid probe corresponding to *vif* protein of HIV-1 would not detect the presence of HIV-2 in a hybridization assay.

The nucleotide sequence of ORF-1 corresponds to *vpr* protein, also known as *R* protein, of HIV-1. The *vpr* protein is not found in HTLV-I or -II (Gallo et al. at 504), and therefore, a nucleic acid probe corresponding to this protein would not detect these viruses in a hybridization assay.

Furthermore, although a *vpr* protein is present in the genome of HIV-2, the nucleotide sequences of the *vpr* proteins of HIV-1 and -2 have a homology of only about 39%. This is shown by a comparison of the nucleotide sequences of the vpr (ORF-1) of HIV-1 given in

LAW OFFICES
FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER, L.L.P.
1300 I STREET, N. W.
WASHINGTON, DC 20005
202-408-4000

applicants' specification with the nucleotide sequence of vpr (ORF-R) of HIV-2 given in Guyader et al., cited above. Exhibit 6 shows the comparison of the two nucleotide sequences. Because there is a homology of only about 39% between the nucleotide sequences of the two proteins, a nucleic acid corresponding to *vpr* protein of HIV-1 would not detect the presence of HIV-2 in a hybridization assay.

The nucleotide sequence of ORF-2 corresponds to *tat* (transactivator) protein, also known as *tat*-3 or TA protein, of HIV-1. The *tat* protein is not found in HTLV-I or -II (Gallo et al. at 504), and therefore, a nucleic acid corresponding to this protein would not detect these viruses a hybridization assay.

Furthermore, although a *tat* protein is present in the genome of HIV-2, the nucleotide sequence of the first exon of the *tat* proteins of HIV-1 and -2 have a homology of only about 48%, and there is almost no homology between the second exon of the *tat* proteins of HIV-1 and -2. This is shown by a comparison of the nucleotide sequences encoding the first exon of the *tat* protein of HIV-1 and HIV-2 (Exhibit 7). The nucleotide sequence of *tat* protein of HIV-1 (ORF-2) is given in applicants' specification, and the nucleotide sequence of *tat* protein of HIV-2 is given in Guyader et al., cited above. For the sequence of *tat* protein of HIV-1, see also Arya et al., "Three Novel Genes of Human T-lymphotropic Virus Type III: Immune Reactivity of Their Products with Sera from Acquired Immune Deficiency Syndrome Patients," Proc. Natl. Acad. Sci., USA, 83, 2209-2213 (1986) (Exhibit 8). Because of the minimal homology between the nucleotide sequences encoding the two proteins, a nucleic acid corresponding to *tat* protein of HIV-1 would not detect the presence of HIV-2 in a hybridization assay.

LAW OFFICES
FINNEGAN, HENDERSON,
FARABOW, GARRETT
& DUNNER, L.L.P.
1300 I STREET, N. W.
WASHINGTON, DC 20005
202-408-4000

The nucleotide sequence of ORF-4 corresponds to *vpu* protein of HIV-1. (See e.g., Cohen et al., "Identification of a Protein Encoded by the vpu Gene of HIV-1," Nature, 334, 532-534 (1988) (Exhibit 9).) This reference gives the amino acid sequence of a protein encoded by the *vpu* gene at page 533, Fig. 1b. A comparison of the nucleotide sequence of ORF-4 and this amino acid sequence reveals that the protein of the reference and applicants' nucleic acid correspond to the same region of the HIV-1 genome.

The *vpu* gene is not found in HTLV-I or -II (Gallo et al. At 504), or in HIV-2 (Cohen et al. At 534, col. 1). Accordingly, a nucleic acid corresponding to the *vpu* gene of HIV-1, when used as a probe in a hybridization assay, would not detect the presence of HTLV-I, HTLV-II, or HIV-2.

Finally, ORF-3 corresponds to nucleotides 5383-5616 and ORF-5 corresponds to nucleotides 7966-8279 of the HIV-1 genome. (Specification at page 13, lines 3 and 5.) ORF-3 is located between the end of the *pol* and *Q* proteins and the beginning of the *env* protein of HIV-1. (See Wain-Hobson et al., "Nucleotide Sequence of the AIDS Virus, LAV," Cell, 40, 9-17 (1985). (Exhibit 10).) Applicants' ORF-3 nucleic acid corresponds to nucleotides 5459-5692 shown at a page 11 of this reference. ORF-5 is located at the end of the *env* protein of HIV-1. Applicants' ORF-5 nucleic acid corresponds to nucleotides 8042-8354 shown at page 12 of Wain-Hobson et al. Corresponding regions are not found in HTLV-I, HTLV-II, or HIV-2. (See the nucleotide sequence of HIV-2 given in Guyader et al.; the nucleotide sequence of HTLV-1 given in Seiki et al., "Human Adult T-cell Leukemia Virus: Complete Nucleotide Sequence of the Provirus Genome Integrated in Leukemia Cell DNA," Proc. Natl. Acad. Sci., USA, 80, 3618-

LAW OFFICES
FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER, L.L.P.
1300 I STREET, N. W.
WASHINGTON, DC 20005
202-408-4000

3622(1983) (Exhibit 11); the nucleotide sequence of 3' region of HTLV-I and -II given in Haseltine et al., "Structure of 3' Terminal Region of Type II Human T Lymphotropic Virus: Evidence for New Coding Region," Science, 225, 419-421, 420 (1984) (Exhibit 12); and the nucleotide sequence of the 3' region of HTLV-I and -II given in Shimotohno et al., "Nucleotide Sequence of the 3' Region of an Infectious Human T-cell Leukemia Virus Type II Genome," Proc. Natl. Acad. Sci., USA, 81, 6657-6661, 6659 (1984) (Exhibit 13).) Accordingly, a nucleic acid corresponding to ORF-3 or ORF-5 of HIV-1, when used as a probe in a hybridization assay, would not detect the presence of HTLV-I, HTLV-II, or HIV-2.

Based on the foregoing remarks and exhibits, it is clear that the peptides recited in the claims are useful to discriminate between retroviruses in diagnostic assays.

In addition, applicants submit that "The enablement analysis should be based on whether there is evidence that one skilled in the art could not have used the compound for any disclosed or well-established use [without] undue experimentation." (35 U.S.C. § 112, First Paragraph, Enablement Training Manual, August 1996, at 21-22.) Therefore, the Examiner must provide evidence that the claimed peptides could not have been used, for example, in hybridization assays. No such evidence has been presented. Therefore, a *prima facie* case of lack of enablement has not been made.

In view of the foregoing remarks, the claimed invention is clearly enabled by the specification and withdrawal of the instant rejection is respectfully requested.

LAW OFFICES
FINNEGAN, HENDERSON,
FARABOW, GARRETT
& DUNNER, L.L.P.
1300 I STREET, N.W.
WASHINGTON, DC 20005
202-408-4000

If there are any other fees due in connection with the filing of this response, please charge the fees to our Deposit Account No. 06-0916. If a fee is required for an extension of time under 37 C.F.R. § 1.136 not accounted for above, such an extension is requested and the fee should also be charged to our Deposit Account.

Respectfully submitted,

FINNEGAN, HENDERSON, FARABOW,
GARRETT & DUNNER, L.L.P.

By: _____
Kenneth J. Meyers
Reg. No. 25,146

ERNEST F. CHAPMAN
Reg. No. 25,961

Dated: December 24, 1996

LAW OFFICES
FINNEGAN, HENDERSON,
FARABOW, GARRETT
& DUNNER, L.L.P.
1300 I STREET, N. W.
WASHINGTON, DC 20005
202-408-4000

Science, 225:927-930 (1984)

Recor

31 August 1984

48. Homology of Genome of AIDS-Associated Virus with Genomes of Human T-Cell Leukemia Viruses

Suresh K. Arya, Robert C. Gallo, Beatrice H. Hahn, George M. Shaw, Mikulas Popovic, S. Zaki Salahuddin, Flossie Wong-Staal

Human T-cell leukemia virus (HTLV) was first identified as an infectious agent etiologically associated with adult T-cell leukemia (ATL) (*1*). A related but distinct retrovirus was isolated from a T-cell variant of hairy cell leukemia (*2*). These viruses, known, respectively, as HTLV-I and HTLV-II, show a tropism for human T cells, particularly OKT4$^+$ cells, and have the capacity to immortalize and transform normal T cells in culture (*3*), alter certain T-cell immune

functions in vitro (4), induce the formation of giant multinucleated T cells (5), and, in some cases, selectively kill certain T cells (6). These properties and data from epidemiologic studies of the acquired immune deficiency syndrome (AIDS), which is uniformly associated with OKT4$^+$ helper cell depletion (7), led us and others to speculate (8) that a member of the HTLV family might be the etiological agent of this disease. In support of this hypothesis was the finding that up to 80 percent of AIDS patients, but less than 1 percent of non-AIDS patients from similar risk groups, have serum antibodies that react with the envelope protein of HTLV (9). However, actual isolations of the known subgroups of HTLV (that is, HTLV-I and HTLV-II) from AIDS patients were infrequent (10).

Recently, we reported repeated isolations of a T lymphotropic retrovirus with cytopathic but not immortalizing activity from patients with AIDS (11). This virus can be grown in a previously immortalized T-cell line (HT) that is relatively resistant to the cytopathic effects of the virus and can grow in the absence of T-cell growth factor (interleukin-2) (12). Using the infected cells as well as purified virus particles in immunological assays, we found that the serum of 80 to 100 percent of AIDS patients and 70 to 80 percent of patients with lymphadenopathy syndrome reacted positively (13). On the basis of its T-cell tropism, the size and Mg^{2+} preference of its reverse transcriptase, the size of its major core protein (24,000 daltons) (14), some antigenic cross-reactivity of its proteins with HTLV-I and HTLV-II (14), and its capacity to induce formation of giant multinucleated cells (12), we considered this virus to be a member of the HTLV family and designated it HTLV-III. Here we show that certain sequences of the genome of HTLV-III and both HTLV-I and HTLV-II are homologous, with the most conserved sequences being located within the *gag-pol* region and less but detectable homology occurring in the *env* and pX region.

Virus particles were purified from supernatant fluids of HT cells, clone 9 (H9) infected with HTLV-III (HTLV-III_B) by centrifugation through a sucrose density gradient at equilibrium (12). HTLV-III_B was originally obtained from pooled supernatants of short-term lymphocyte cultures of AIDS patients. Virus particles were also purified from normal peripheral blood lymphocytes newly infected by virus of a primary leukocyte culture of another AIDS patient (HTLV-III_Z) (11). The particles were lysed with sodium dodecyl sulfate (SDS), digested with proteinase K, and directly chromatographed on an oligo(dT) cellulose column. The resulting polyadenylate [poly(A)]-containing RNA was used as template to synthesize ^{32}P-labeled complementary DNA (cDNA) in the presence of oligo(dT) primers. The size of the resultant cDNA ranged from 0.1 to 10 kb (not shown). When these labeled cDNA's were hybridized to poly(A)-containing RNA purified from infected and uninfected H9 cells as well as other uninfected human cell lines, only the infected H9 cells contained homologous RNA sequences as evidenced by discrete RNA bands after Northern hybridization. Figure 1 shows that cDNA preparations from HTLV-III_B and HTLV-III_Z gave identical patterns, detecting RNA species of about 9.0, 4.2, and 2.0 kb. These bands are similar in size to those corresponding to genomic size messenger RNA (mRNA) and spliced mRNA's of *env* and pX sequences previously observed in cells infected with HTLV-I

(15), consistent with the anticipated relatedness of these viruses. Furthermore, viral mRNA bands of HTLV-II–infected cells were detected with an HTLV-III cDNA probe (Fig. 1b, lane 6) and again the sizes of the mRNA were like those with HTLV-I.

To determine directly the homology between HTLV-III and HTLV-I and HTLV-II, we hybridized HTLV-III cDNA to cloned genomes of HTLV-I and HTLV-II digested with specific restriction endonucleases. Complete genomes of a prototype HTLV-I (16), an HTLV-I variant called HTLV-Ib (16), and HTLV-II were digested with two restriction enzymes as indicated in the legend to Fig. 2 and blot-hybridized to ^{32}P-labeled HTLV-III$_B$ cDNA. A region spanning the *gag* and *pol* genes showed the greatest homology. For the prototype HTLV-I, this corresponds to the 1.7-kb Pst I–Pst I fragment and 5.3-kb Sst I–Sal I fragment. HTLV-Ib, which lacks a Pst I site indicated in parentheses in Fig. 2, revealed the expected 3.0-kb Pst I–Pst I fragment instead. Similarly, strong hybridization to the *gag-pol* sequences of HTLV-II also occurred. This is reflected in the 4.2-kb Bam HI–Xho I fragment and the 4.0-kb Bam HI–Eco RI fragment (Fig. 2, lanes 5 and 6).

Fragments corresponding to the *env* and pX sequences of HTLV-I and HTLV-II also hybridized weakly with HTLV-III$_B$ cDNA (see the 2.4-kb Pst I–Pst I and the 2.1-kb Sst I–Pst I fragment in Fig. 2, lane 1) as did the 1.4-kb Pst I fragment of HTLV-Ib containing only pX sequences (Fig. 2, lane 4). The ease of detection of these sequences varied with different preparations of cDNA, probably because of variable representations of the 3′ end of the virus genome. We used cDNA from both HTLV-III$_B$ and HTLV-III$_Z$. Figure 3 shows the results for HTLV-III$_Z$ cDNA. Subclones of HTLV-I containing different regions of the genome were hybridized to

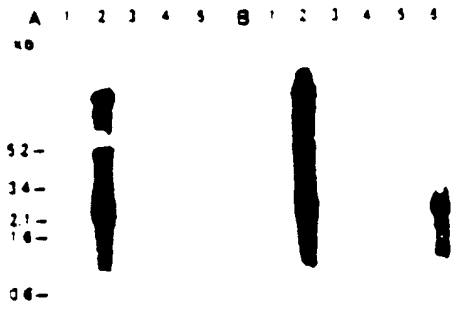


Fig. 1. HTLV-III–specific sequences in cellular RNA from HTLV-infected cells. Poly(A)-selected cellular RNA was size-separated by formaldehyde-agarose gel electrophoresis, transferred to Zeta probe membrane (Bio-Rad Labs) by electroelution and hybridized to (A) HTLV-III$_B$ cDNA and (B) HTLV-III$_Z$ cDNA. (A and B) Lane 1, uninfected H9 cells (5 μg); lane 2, HTLV-III$_B$–infected H9 cells (10 μg); lane 3, leukemic Jurkat cells (10 μg); lane 4, HTLV-I–infected C5/MJ cells (5 μg); and lane 5, HTLV-II–infected MO cells (5 μg). (B) Lane 6, a longer exposure of lane 5 in (B). Poly(A)-selected RNA was prepared by guanidine-HCl extraction and cesium chloride centrifugation followed by oligo(dT) cellulose chromatography as described (24). The cDNA was transcribed from poly(A)-selected virus-associated RNA with the use of oligo(dT) as a primer and avian myeloblastosis virus RNA-directed DNA polymerase as described (25). The hybridization was performed at 37°C for 16 hours in a mixture containing 40 percent formamide, 5× standard sodium chloride and sodium citrate (SSC; 0.15*M* NaCl and 0.015*M* sodium citrate, pH 7), 0.05*M* sodium phosphate buffer (pH 7), 5× PM (0.02 percent each of bovine serum albumin, polyvinylpyrrolidone, and Ficoll 400), yeast RNA (200 μg/ml), denatured salmon sperm DNA (20 μg/ml), 0.1 percent SDS, and 10 percent dextran sulfate. The membrane was subsequently repeatedly washed with 2× SSC and 0.1 percent SDS at 62°C, air-dried, and exposed to a Kodak XAR film with the use of intensifying screens.

HTLV-III$_Z$ cDNA (Fig. 3A). With the exception of fragment c, which corresponds to an internal portion of the *pol* gene, all fragments were detected by hybridization, including fragment a (LTR-*gag*) after long exposure of the

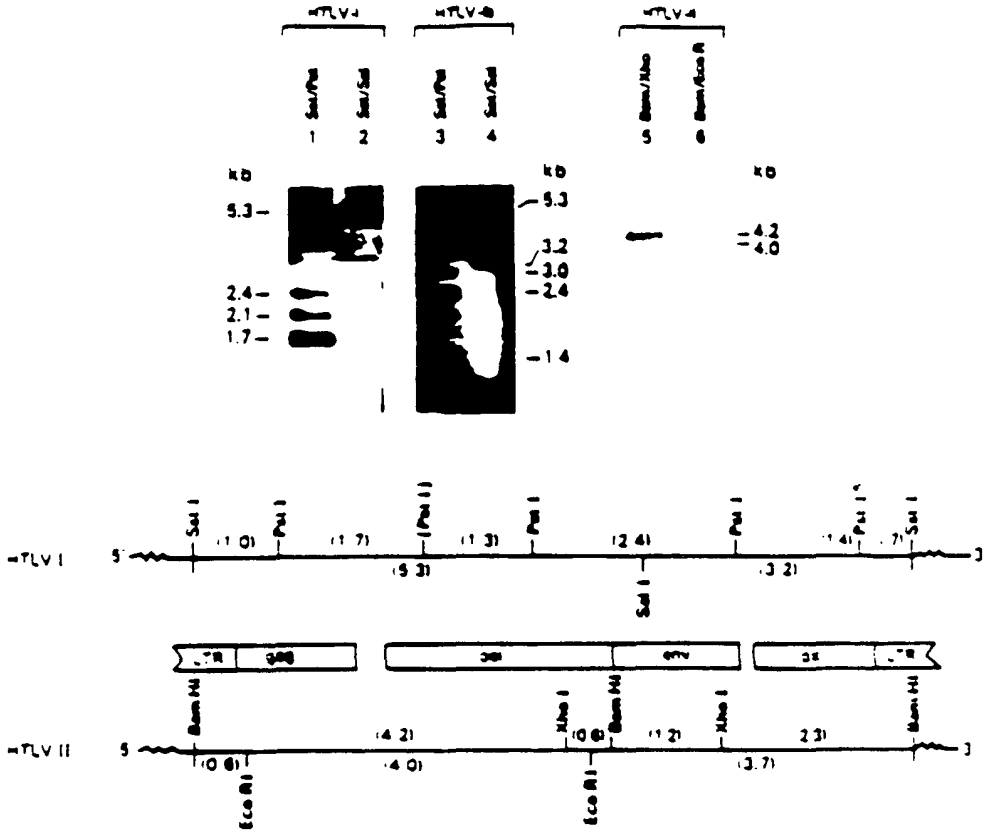


Fig. 2. Relatedness of the genome of HTLV-III$_B$ with the genomes of HTLV-I and HTLV-II. Sites of digestion by the relevant restriction enzymes and the expected sizes of the fragments are shown below the gels. Cloned HTLV-I (λST), HTLV-Ib (λMC), and HTLV-II (pMO) DNA's were digested with the indicated restriction enzymes and fragments were separated by agarose gel electrophoresis, transferred to a nitrocellulose membrane (*23*), and hybridized with HTLV-III$_B$cDNA. Lanes 1 and 2, HTLV-I (λST) DNA digested with Sst I plus Pst I and Sst I plus Sal I, respectively; lanes 3 and 4, HTLV-Ib (λMC) DNA digested with Sst I plus Pst I and Sst I plus Sal I, respectively; lanes 5 and 6, HTLV-II (pMO) DNA digested with Bam HI plus Xho I and Bam HI plus Eco RI, respectively. HTLV-I (λST) and HTLV-I (λMC) clones were obtained from the genomic libraries of DNA's from ATL patients S.T. and M.C., respectively. Both cellular DNA's were cloned at the Sst I site of phage λgtWES · λB DNA (*16*). HTLV-I (λST) is a prototype HTLV-I and HTLV-Ib (λMC) is a variant of HTLV-I that contains some divergent restriction enzyme sites, including the lack of the second Pst I site from the 5' end of the viral genome (*16*). HTLV-II (pMO) was obtained by subcloning λMO15A (*26*) at the Bam HI site of plasmid pBR322 DNA. The cDNA was synthesized as described in Fig. 1 and hybridization was performed at 37°C for 16 hours in a mixture containing 30 percent formamide, 5× SSC, 5× PM, denatured DNA (100 μg/ml), 0.1 percent SDS, and 10 percent dextran sulfate. The membrane was subsequently washed and exposed as described in Fig. 1.

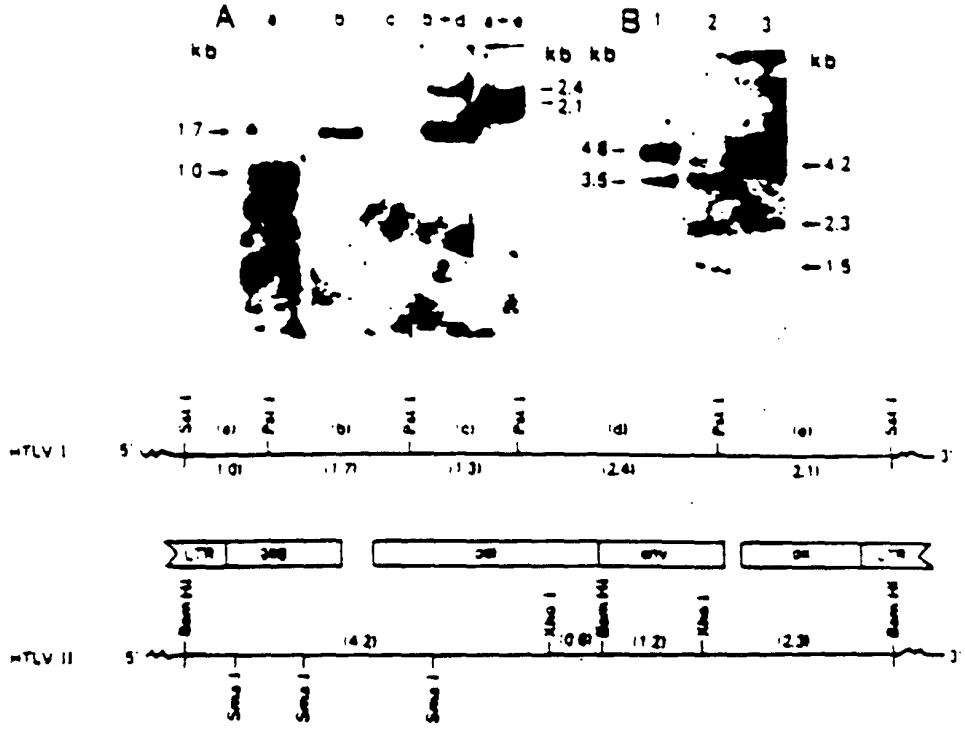


Fig. 3. Relatedness of the genome of HTLV-III$_Z$ with the genomes of HTLV-I and HTLV-II. DNA from subclones of HTLV-I$_{ST}$ and HTLV-II$_{MO}$ was digested with the indicated restriction enzymes. Fragments were separated by agarose gel electrophoresis, transferred to a nitrocellulose membrane (24), and hybridized with HTLV-III$_Z$ cDNA. (A) HTLV-I subclones were constructed by "shotgun" cloning of fragments generated by codigestion with Pst I and Sst I into pBR322 containing fragments designated a to e on the illustrated restriction map of HTLV-I. The viral inserts were released by digestion with the appropriate enzymes. (B) HTLV-II (pMO) DNA: Lane 1, digested with Bam HI; lane 2, digested with Bam HI plus Sma I; lane 3, digested with Bam HI plus Xho I. The cDNA was synthesized as in Fig. 1 and hybridization was performed as in Fig. 2, except that the hybridization mixture contained 40 percent formamide.

autoradiogram. Similarly, the 3' half of HTLV-II contained in the 3.5-kb Bam HI–Bam HI fragment and the 2.3-kb Bam HI–Xho I fragment could be detected with this particular HTLV-III cDNA probe (Fig. 3B).

Retroviruses called LAV (or sometimes IDAV$_1$ and IDAV$_2$) have been isolated from patients with lymphadenopathy syndrome and AIDS (17). Although LAV has been reported to lack relatedness to HTLV-I and -II (17), further characterization of its proteins and nucleic acids may reveal that LAV is related to these viruses and is identical to or related to HTLV-III.

The present data showing that certain nucleotide sequences of HTLV-III are homologous to sequences of HTLV-I and HTLV-II support our proposal that this virus should be classified within the HTLV family. However, HTLV-III is much less related to HTLV-II and HTLV-I than HTLV-II and HTLV-I are to each other. It is of interest that still other HTLV-related T lymphotropic ret-

roviruses have been identified in Old World monkeys (*18*). These primate viruses are closely related to HTLV-I and only minimally to HTLV-II (*19*). Although the most conserved sequences of HTLV-III are in the region spanning the junction of the predicted *gag* and *pol* genes, other weakly homologous sequences are also detected in the *env* and pX genes. Homology in the *gag* and *env* coding sequences has already been suggested by immunological cross-reactivity between these antigens derived from the three subgroups (*14*). Homology in the pX region is an additional demonstration that HTLV-III belongs to the HTLV family, which is unique among retroviruses in its possession of the pX genes (*20, 21*). It is interesting that pX is the most conserved region between HTLV-I and HTLV-II (*21*) and that both of these viruses can transform T cells in vitro. In contrast, the pX region is much less conserved in HTLV-III, a cytopathic virus that lacks transforming activity (*11, 12*).

Comparisons of the LTR regions between HTLV-I and HTLV-II have revealed a conserved 21-bp repeat sequence in two otherwise very divergent LTR's (*22*). The location of this sequence upstream of promoter sequences suggests that it is similar to other viral enhancer sequences. In view of the tropism of HTLV-III for OKT4$^+$ lymphocytes, it will be interesting to see if this virus also has such an enhancer sequence in its LTR. Our present study does not allow us to compare specifically the LTR of HTLV-III to those of HTLV-I and -II. However, the weak signal obtained with 5' and 3' ultimate fragments containing the LTR suggest that these elements have minimal or no homology.

References and Notes

1. R. C. Gallo and F. Wong-Staal, *Blood* **60**, 545 (1982); R. C. Gallo, in *Cancer Surveys*, L. M. Franks, I. M. Wyke, R. A. Weiss, Eds. (Oxford Univ. Press, Oxford, 1984), vol. 3, pp. 113–159.
2. V. S. Kalyanaraman *et al.*, *Science* **218**, 571 (1982).
3. I. Miyoshi *et al.*, *Nature (London)* **294**, 770 (1981); M. Popovic, G. Lange-Wantzin, P. S. Sarin, D. Mann, R. C. Gallo, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 5402 (1983); P. D. Markham *et al.*, *Int. J. Cancer* **31**, 413 (1983); I. S. Y. Chen *et al.*, *Nature (London)* **308**, 502 (1983); M. Popovic *et al.*, in *Human T-Cell Leukemia Viruses*, R. C. Gallo, M. Essex, L. Gross, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., in press).
4. M. Popovic *et al.*, in preparation.
5. M. Popovic, F. Wong-Staal, P. S. Sarin, R. C. Gallo, *Adv. Viral Oncol.* **4**, 45 (1984); K. Nagy, P. Clapham, R. Cheingsong-Popov, R. A. Weiss, *Int. J. Cancer* **32**, 321 (1983).
6. H. Mitsuya *et al.*, *Science* **223**, 1293 (1984).
7. Centers for Disease Control Task Force on Kaposi's Sarcoma and Opportunistic Infections, *N. Engl. J. Med.* **306**, 248 (1982); J. P. Haaraan, G. P. Wormser, C. P. Macguire, L. J. DeLorenzo, G. Davis, *ibid.* **307**, 498 (1982); J. W. Curran *et al.*, *ibid.* **310**, 69 (1984).
8. R. C. Gallo, P. S. Sarin, W. A. Blattner, F. Wong-Staal, M. Popovic, in *Seminars in Oncology: AIDS*, J. E. Groopman, Ed. (Grune & Stratton, San Diego, 1984), pp. 12–17; M. Essex *et al.*, in *Human T-Cell Leukemia Viruses*, R. C. Gallo, M. Essex, L. Gross, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., in press).
9. M. Essex *et al.*, *Science* **220**, 859 (1983); *ibid.* **221**, 1061 (1983); T. H. Lee *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
10. R. C. Gallo *et al.*, *Science* **220**, 865 (1983); B. H. Hahn *et al.*, in *Acquired Immune Deficiency Syndrome, UCLA Symposia on Molecular and Cellular Biology*, M. S. Gottlieb and J. E. Groopman, Eds. (Liss, New York, in press).
11. R. C. Gallo *et al.*, *Science* **224**, 500 (1984).
12. M. Popovic, M. G. Sarngadharan, E. Read, R. C. Gallo, *ibid.*, p. 497.
13. M. G. Sarngadharan, M. Popovic, L. Bruch, J. Schupbach, R. C. Gallo, *ibid.*, p. 506; M. G. Sarngadharan *et al.*, in preparation.
14. J. Schupbach *et al.*, *Science* **224**, 503 (1984).
15. G. Franchini, F. Wong-Staal, R. C. Gallo, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
16. B. Hahn *et al.*, *Int. J. Cancer*, in press.
17. F. Barre-Sinoussi *et al.*, *Science* **220**, 868 (1983); E. Vilmer *et al.*, *Lancet* **1984-I**, 753 (1984); L. Montagnier, personal communication.
18. I. Miyoshi *et al.*, *Gann* **74**, 323 (1983); C. W. Sasseger *et al.*, in *Human T-Cell Leukemia Viruses*, R. C. Gallo, M. Essex, L. Gross, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., in press).
19. H.-G. Guo, F. Wong-Staal, R. C. Gallo, *Science* **223**, 1195 (1984).
20. M. Seiki, S. Hattori, Y. Hirayama, M. Yoshida, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3618 (1983).
21. G. M. Shaw *et al.*, *ibid.* **81**, 4544 (1984).
22. J. Sodroski *et al.*, *ibid.*, in press.

23. E. M. Southern, *J. Mol. Biol.* **98**, 503 (1973).
24. R. A. Cox, *Methods Enzymol.* **12**, 120 (1967); S. L. Adams *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3399 (1980).
25. T. Maniatis, E. F. Fritsch, J. Sambrook, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1983), pp. 233–233.
26. E. P. Geiduschek *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 993 (1984).
27. We gratefully acknowledge the help and advice of Dr. P. Markham and the expert editorial assistance of A. Mozzuca.

1 May 1984; accepted 30 May 1984

Z

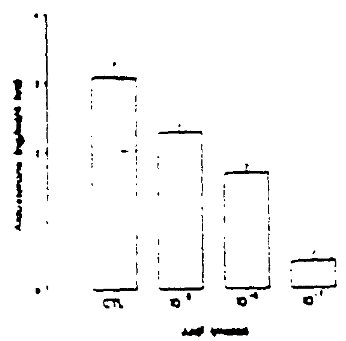


Fig. 1 Effect of ANF(8-33) on basal aldosterone secretion. Rat glomerulosa cells were prepared by enzymatic digestion of 20 rat adrenals after enucleation. The cells remaining on the capsule were digested for 30 min with a mixture of collagenase and DNase (4 mg ml^{-1}, 4 μg ml^{-1}) for 30 min. Dispersed cells were filtered through gauze and centrifuged at 800 r.p.m. for 15 min. The pellet was resuspended in M199 buffer containing 0.1% bovine serum albumin (BSA) and the cells centrifuged at 800 r.p.m. for 15 min. The cell pellet was again resuspended in M199-0.1% BSA buffer and distributed in 900-μl aliquots to 12 × 75 plastic tubes. The samples were preincubated for 90 min in a 37 °C waterbath under an atmosphere of 5% CO_2/95% O_2. Aliquots of the test samples were added in a 100 μl volume and incubated for 4 h. Aldosterone and corticosterone were measured by radioimmunoassay using antisera purchased from Endocrine Sciences, Oxnard, California, and ^{3}H-labelled steroid from NEN. Results are the mean ± s.e.m. of seven replicates. Statistical analysis was performed by analysis of variance and all points are significant ($P < 0.01$) from control. (ANF(8-33) was the gift of Drs R. Hirschmann and D. F. Veber of Merck, Sharp and Dohme Research Laboratories).

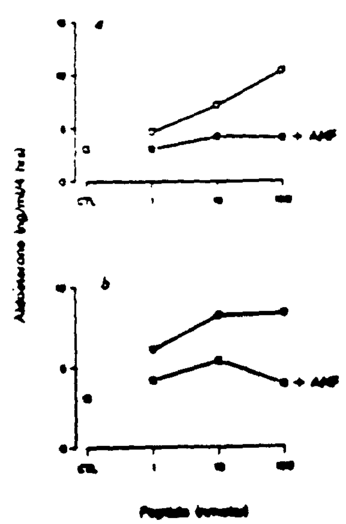


Fig. 2 Effect of ANF(8-33) on stimulated aldosterone secretion. a, Rat glomerulosa cells were prepared as described in Fig. 1 and incubated with synthetic human ACTH either alone (open circles) or in combination with equimolar amounts of ANF(8-33) (closed squares). Aldosterone secretion was measured as above by radioimmunoassay. b, Cells were incubated with synthetic angiotensin-II (AN-II) either alone (open circles) or in the presence of equimolar amounts of ANF(8-33) (closed squares). Control cells received neither peptide, thereby indicating the ability of ANF to decrease aldosterone production to basal levels. Results are the mean ± s.e.m. of 7 replicates and all points are significant when compared with their respective control ($P < 0.001$). Synthetic hACTH(1-39) and angiotensin II were synthesized by Dr Nicholas Ling by solid-phase methodology.

(8-33) as a natriuretic hormone, and now in inhibiting basal and stimulated aldosterone formation, suggests that its biological activities are an integral part of the homeostatic mechanisms regulating sodium retention. Furthermore, unlike somatostatin, its inhibitory effect is not restricted to angiotensin-stimulated aldosterone secretion, but affects the formation of both basal and stimulated mineralocorticoids. Moreover, at no point was ANF(8-33) observed to stimulate aldosterone. The observations reported here provide the groundwork for defining the mechanisms by which atrial-derived peptides affect sodium retention and suggest that this peptide may be responsible for the attenuated effects of AN-II on the adrenal cortex during sodium loading[16,17]. The understanding of some clinical forms of idiopathic hypo- and hypertension[18,19] may therefore result from defining the interactions between ANF, the adrenal cortex and the basic mechanisms regulating ANF secretion.

After submission of this manuscript, Chartier *et al.*[20] and DeLean *et al.*[21] reported findings similar to those reported here.

We thank Drs R. Guillemin and P. Böhlen for their critical review and comments on this manuscript and the secretarial staff of the Laboratories for Neuroendocrinology for help in preparation of the manuscript. This research was supported by grants from the NIH (HD-09690 and AM-18811) and the Robert J. Kleberg Jr. and Helen C. Kleberg Foundation.

Received 18 May; [illegible] in revised form 16 July; accepted 15 November 1984.

1. Flynn, T. G., de Bold, M. L. & de Bold, A. J. *Biochem. biophys. Res. Commun.* 117, 859–865 (1983).
2. Currie, M. G. *et al. Science* 223, 67–69 (1984).
3. Kangawa, K. & Matsuo, H. *Biochem. biophys. Res. Commun.* 118, 131–139 (1984).
4. Jamieson, J. D. & Palade, G. E. *J. Cell Biol.* 23, 151–171 (1964).
5. Sonnenberg, H., Cupples, W. A., de Bold, A. J. & Veress, A. T. *Can. J. Physiol. Pharmac.* 60, 1149–1152 (1982).
6. Keeler, R. *Can. J. Physiol. Pharmac.* 60, 1078–1082 (1982).
7. Thibault, G. *et al. FEBS Lett.* 167, 352–356 (1984).
8. [illegible], S. A. S. *et al. Biochemistry* 8, 775–787 (1974).
9. [illegible] *et al. Clin. Endocr.* 12, 45–50 (1980).
10. Douglas, J., Aguilera, G., Kondo, T. & Catt, K. J. *Endocrinology* 102, 685–696 (1978).
11. Hornsby, P. J. & Gill, G. N. *Endocrinology* 102, 926–936 (1978).
12. Aguilera, G., Harwood, J. P. & Catt, K. J. *Nature* 292, 262–263 (1981).
13. [illegible] (1982).
14. [illegible] *Life Sci.* 28, 1355–1361 (1981).
15. [illegible] *et al. J. clin. Endocr. Metab.* 54, 636–640 (1982).
16. [illegible] *et al. Endocrinology* 102, [illegible] (1978).
17. Cowley, A. W. & McCaa, R. E. *Circulation Res.* 39, 788–797 (1976).
18. [illegible], M. I. & Levine, L. S. *Endocrine Rev.* 1, 421–430 (1980).
19. [illegible] *et al. J. clin. Endocr. Metab.* 53, 867–873 (1981).
20. Chartier, L., Schiffrin, E., Thibault, G. & Garcia, R. *Endocrinology* (in the press).
21. DeLean, A. *et al. Endocrinology* (in the press).

Molecular cloning of lymphadenopathy-associated virus

Marc Alizon*§, Pierre Sonigo†,
Françoise Barré-Sinoussi*, Jean-Claude Chermann*,
Pierre Tiollais†, Luc Montagnier*
& Simon Wain-Hobson†‡

*Unité d'Oncologie Virale et Equipe de Recherche CNRS 147, and
†Unité de Recombinaison et Expression Génétique (INSERM U 163, CNRS LA 271), Institut Pasteur, 25 rue du Dr Roux,
75724 Paris Cédex 15, France

Lymphadenopathy-associated virus (LAV) is a human retrovirus first isolated[1] from a homosexual patient with lymphadenopathy syndrome, frequently a prodrome or a benign form of acquired immune deficiency syndrome (AIDS)[2]. Other LAV isolates have subsequently been recovered from patients with AIDS or pre-AIDS[3-4] and all available data are consistent with the virus being the causative agent of AIDS. The virus is propagated on activated T lymphocytes and has a tropism for the T-cell subset OKT4 (ref.

‡ To whom correspondence should be addressed.
§ On leave from: Division de Radiobiologie et Radioprotection, CRSSA, 1 rue du Lt R. Batany, 92140 Clamart, France.

6), in which it induces a cytopathic effect. The major core protein of LAV is antigenically correlated to other known retroviral antigens[1,2,7]. LAV-like viruses have more recently been independently isolated from patients with AIDS and pre-AIDS. These viruses, called human T-cell leukaemia/lymphoma virus type III (HTLV-III)[8–11] and AIDS-associated retrovirus (ARV)[12], seem to have many characteristics in common with LAV and probably represent independent isolates of the LAV prototype. We have sought to characterize LAV by the molecular cloning of its genome. A cloned LAV complementary DNA was used to screen a library of recombinant phages constructed from the genomic DNA of LAV-infected T lymphocytes. Two families of clones were characterized which differ in a restriction site. The viral genome is longer than any other human retroviral genome (9.1–9.2 kilobases).

The cDNA first-strand of LAV was synthesized in an endogenous, detergent-activated reaction. LAV virions were purified from the supernatant of FRA cells, a B-lymphoblastoid LAV-producing line[13], and the reaction was primed with oligo(dT). Three cDNA clones, pLAV13, 75 and 82, carrying inserts of 2.5, 0.6 and 0.8 kilobases (kb), respectively, were characterized further (Fig. 1). All three inserts have a common restriction pattern at one end, indicative of a common priming site. The 50-base pair (bp) common *Hin*dIII–*Pst*I fragment was sequenced and shown to contain an oligo(dA) stretch preceding

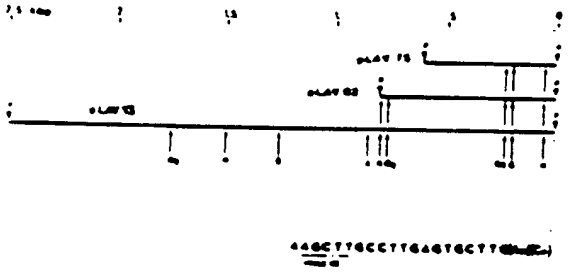


Fig. 1 Restriction maps of cDNA clones derived from LAV genomic RNA. Restriction sites: B, *Bam*HI; Bg, *Bgl*II; H, *Hin*dIII; K, *Kpn*I; S, *Sac*I; X, *Xho*I.
Methods: LAV cDNA was synthesized in an endogenous detergent-activated reaction. For each reaction, LAV virions were purified on a 20–60% sucrose gradient as described previously[1], from 200 ml of supernatant of the LAV-producing FRA line[13]. Virus-containing fractions were pooled, diluted with NTE buffer (100 mM NaCl, 10 mM Tris-HCl pH 7.8, 1 mM EDTA) and centrifuged (Beckman type SW56 rotor, 50,000 r.p.m., 60 min). The viral pellet was resuspended in 250 μl of NTE. Reaction volume was adjusted to 1 ml and final concentrations were: 50 mM Tris-HCl pH 7.8, 25 mM NaCl, 6 mM $MgCl_2$, 10 mM dithiothreitol, 0.02% Triton X-100, 0.1 mM of each of dATP, dGTP, TTP, 4 μM dCTP including 200 μCi of [α-^{32}P]dCTP (400 Ci mmol^{-1}, Amersham) and 50 μg ml^{-1} oligo(dT) primer. Incubation was at 37 °C. After 15 min, dCTP was added to 25 μM. At 45 min, the reaction was stopped with EDTA and SDS (final concentrations 20 mM and 0.5%, respectively). After 1 h of proteinase K digestion (100 μg ml^{-1}, 37 °C), the reaction mixture was extracted with phenol/chloroform and cDNA–RNA hybrids were ethanol-precipitated. Second-strand synthesis with exonuclease-free DNA polymerase I (Boehringer) and RNase H (BRL) and dC-tailing with terminal transferase (Boehringer) were performed according to Gubler and Hoffman[20]. Tailed double-stranded cDNA was annealed to dG-tailed *Pst*I-linearized pBR327 vector. *Escherichia coli* C600 *rec*BC was transformed by the $CaCl_2$ method; 500 recombinant clones were screened *in situ*[27] with a ^{32}P-labelled LAV cDNA in which the first strand had been synthesized as described above, except that an alkaline hydrolysis step was included. Approximately 10% of recombinants proved positive, the majority of which formed a family of cross-hybridizing clones. Three recombinants, pLAV13, pLAV75 and pLAV82, carrying inserts of 2.5, 0.6 and 0.8 kb, respectively, were analysed further. There are no sites for *Eco*RI, *Nru*I, *Pst*I, *Sal*I, *Sma*I, *Sst*I or *Xho*I in the pLAV13 insert. The *Hin*dIII–*Pst*I fragment was subcloned into M13mp8 and sequenced according to Sanger *et al.*[28] using a 15-mer primer (Bielabs) and [α-^{32}P]dCTP (Amersham).

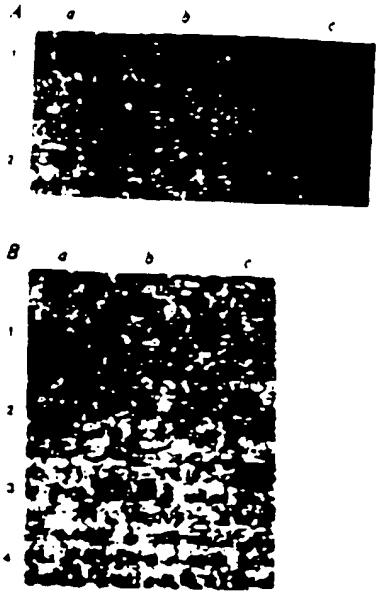


Fig. 2 Rapid dot-blot technique for LAV detection in cell culture supernatant. Spots represent: A, a, 1 μl; b, 2 μl; c, 4 μl of concentrated (250 ×) cell culture supernatant from (1) LAV-producing CEM cells (reverse transcriptase activity (RT), determined as described previously[1], was 140,000 c.p.m. ml^{-1}); (2) LAV-producing Epstein–Barr-transformed B-cell line FRA (RT 175,000 c.p.m. ml^{-1}). B, a, 1 μl; b, 2 μl; c, 5 μl of 100 × concentrated supernatant from (1) uninfected normal T lymphocytes (no RT activity); (2) LAV-producing normal T lymphocytes (RT 170,000 c.p.m.); (3) LAV-producing CEM line (RT 150,000 c.p.m.); and (4) culture of bone marrow lymphocytes from a haemophiliac patient[4] with AIDS (RT 7,000 c.p.m.).
Methods: Cell culture supernatants were pelleted through 0.5 ml 20% sucrose cushions in NTE buffer (Beckman type SW56 rotor, 50,000 r.p.m., 1 h, 4 °C). The pellet was resuspended in NTE buffer as indicated. Concentrated virus was spotted onto dried nylon filters (Zetabind) presoaked in 20 × SSC (3 M NaCl, 0.3 M sodium citrate). After baking (at least 30 min at 80 °C), filters were hybridized with ^{32}P nick-translated pLAV13 insert (Fig. 1) (specific activity > 10^8 c.p.m. per μg) for 12–16 h in stringent conditions (50% formamide, 5 × SSC, 42 °C), washed (0.1 × SSC, 0.1% SDS, 65 °C, 2 × 30 min), and exposed for 20 h (Kodak XAR5 film with an intensifying screen) at −70 °C.

the cloning dC tail. The clones are thus copies of the 3′ end of a poly(A) RNA.

The specificity of pLAV13 was determined in a series of filter hybridization experiments using nick-translated pLAV13 insert as a probe. First, using an adapted spot-blot technique, we could detect LAV virion RNA from normal T cells, FRA and other B-cell lines and CEM cells (L.M. and R. Weiss, unpublished results; Fig. 2). LAV was also detected in a bone marrow cell culture (Fig. 2B, line 4) from a haemophiliac with AIDS[4], in spite of the low titre of virus in the supernatant. Uninfected cultures proved negative (Fig. 2B, line 1). Second, the probe detected DNA in the Southern blots of LAV-infected T lymphocytes and CEM cells (Fig. 3). No hybridization was detected in DNA from uninfected lymphocytes or from normal liver (data not shown) in the same hybridization conditions. A characteristic 1.45-kb *Hin*dIII fragment which co-migrated with an internal viral fragment in *Hin*dIII-cleaved pLAV13 (Fig. 1) was detected in the Southern blots. Bands at 2.3 and 6.7 kb were also detected. Together, these data show that pLAV13 DNA is exogenous to the human genome and detects both RNA and integrated DNA forms derived from LAV-infected cells. Thus, pLAV13 is LAV specific. Being oligo(dT)-primed, pLAV13 must contain the R and U3 regions of the long terminal repeat (LTR) as well as

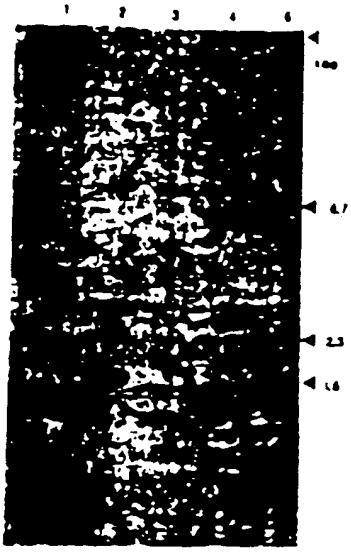


Fig. 3 Southern blot of *Hind*III-restricted genomic DNA from LAV-infected and uninfected cells hybridized with pLAV13. *Hind*III-restricted high relative molecular mass DNA from: lane 1, uninfected CEM cells; lane 2, LAV-infected CEM cells; lane 3, uninfected T cells after 5 days' culture; lane 4, LAV-infected T cells 2 days after infection; lane 5, LAV-infected T cells 5 days after infection.

Methods: Peripheral blood T lymphocytes of a healthy donor were stimulated for 3 days with phytohaemagglutinin, after which they were infected with LAV (isolate BRU-LAV1) at 10^4 c.p.m. reverse transcriptase activity per 10^6 cells as described previously[1], except for part of the culture kept uninfected for controls. Two and five days after infection, genomic DNA was extracted. *Hind*III-digested DNA (10 μg) was electrophoresed through a 0.8% agarose gel and Southern blotted. The filter was hybridized in 10 ml of 50% formamide, 5 × SSC, 1 × Denhardt's, 10% dextran sulphate with 100 μg ml^{-1} denatured sonicated salmon sperm DNA and 2×10^7 c.p.m. of nick-translated pLAV13 insert (4×10^4 c.p.m. per μg) for 10 h at 42 °C. The filter was washed at 68 °C in 0.1 × SSC, 0.1% SDS for 2 × 30 min and exposed to Kodak XAR5 film at −70 °C for 16 h using an intensifying screen.

the 3' end of the coding region, assuming a conventional retroviral genome structure.

Having found a *Hind*III site about 20 bp 5' of the poly(A) stretch and thus within the R region of the LTR, we cloned the LAV genome by making a partial *Hind*III digest of genomic DNA from LAV-infected T cells of a healthy donor. A 9 ± 1.5-kb DNA-containing fraction was precipitated and ligated into the *Hind*III arms of phage vector λL47.1 (ref. 14). When nick-translated pLAV13 insert was used as a probe to screen ~2×10^6 phage plaques *in situ*, five independent clones were obtained. A restriction map of clone λJ19 and of a *Hind*III variant, λJ81, are shown in Fig. 4. Recombinants λJ27, λJ31 and λJ57 have the same *Hind*III map as λJ19, while λJ81 is so far unique. As the two clones were derived from the first isolate[1] of LAV reported (isolate BRU, or LAV1), we refer to the two viral genomes as LAV1a (λJ19) and LAV1b (λJ81). λJ19 shows four *Hind*III bands of 6.7, 1.45, 0.6 and 0.52 kb, the first two of which correspond to bands in the genomic blot of *Hind*III-restricted DNA (Fig. 3, lane 5). The smallest bands (0.6 and 0.52 kb) were not seen in the genomic blot, but the fact that they appear in all the independently derived clones analysed indicates that they represent internal and not junction fragments, assuming random integration of LAV proviral DNA. However, the 0.52-kb band hybridizes with pLAV13 DNA (Fig. 4) through the small *Hind*III–*Pst*I fragment of pLAV13. Thus, the 0.5-kb *Hind*III fragment of λJ19 contains the R/U5 junction within the LTR. The finding of two small *Hind*III fragments in the 5' region reinforces

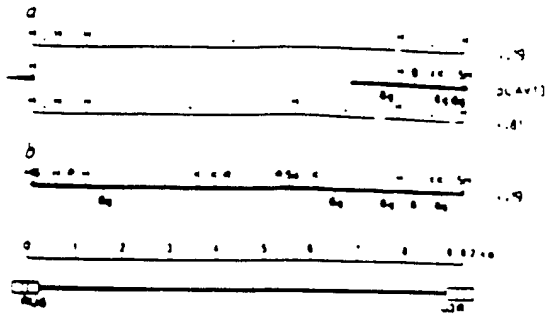


Fig. 4 Restriction maps of LAV proviral DNA in clones λJ19 (LAV1a) and λJ81 (LAV1b). a, *Hind*III restriction maps of LAV proviral DNA in clones λJ19 and λJ81. Those *Hind*III fragments detected by pLAV13 are marked by +, those not, by −. The restriction map of the pLAV13 cDNA clone is also shown. b, Restriction map of λJ19. Restriction sites: B, *Bam*HI; Bg, *Bgl*II; H, *Hind*III; K, *Kpn*I; P, *Pst*I; R, *Eco*RI; S, *Sac*I; Sa, *Sal*I; X, *Xho*I. Beneath the scale is a schema for the general structure of retroviruses showing the LTR elements U3, R and U5. Only the R/U5 boundary has been defined (Fig. 1) and other boundaries are drawn only figuratively.

Methods: DNA from LAV-infected T cells was partially digested with *Hind*III and fractionated on a 5–40% sucrose gradient in 10 mM Tris-HCl pH 8, 10 mM EDTA, 1 M NaCl (Beckman type SW41 rotor, 16 h, 40,000 r.p.m.). A single fraction (9 ± 1.5 kb) was precipitated with 20 μg ml^{-1} dextran T40 as carrier and taken up in TE buffer (10 mM Tris-HCl pH 8, 1 mM EDTA). λL47.1 (ref. 14) *Hind*III arms were prepared by first ligating the *Eco*RI sites followed by *Hind*III digestion and fractionation through a 5–40% sucrose gradient as above. Fractions containing only the λ *Hind*III arms were pooled, precipitated and taken up in TE buffer. Ligation of arms to DNA was made at ~200 μg ml^{-1} DNA using a 3:1 molar excess of arms and 300 U of T4 DNA ligase (Biolabs). *In vitro* packaging lysates were made according to ref. 29. After *in vitro* packaging, the phage lysate was plated out on NM538 or a C600 *rec*BC strain. Approximately 2×10^6 plaques were screened by *in situ* hybridization[30] using nitrocellulose filters. Hybridization was performed at 68 °C in 1 × Denhardt's solution, 0.5% SDS, 2 × SSC, 2 mM EDTA. Probe: ^{32}P nick-translated insert of pLAV13 at $> 10^8$ c.p.m. per μg. Filters were washed for 2 × 30 min in 0.1 × SSC 0.1% SDS at 68 °C, and exposed to Kodak XAR-5 film for 24–40 h with intensifying screens at −70 °C. Seven positive clones were identified and plaque-purified on a C600 *rec*BC strain. Liquid cultures were grown and the recombinant phages banded in CsCl. Phage DNA was extracted and digested in the appropriate conditions. The restriction maps were orientated by hybridizing blots to pLAV13 DNA, which maps the 3' coding sequences of the viral genome as well as the U3–R region of the LTR. All cloning and amplification of LAV genomic clones was carried out in a P3 laboratory.

the usefulness of cloning LAV by partial restriction of genomic DNA.

λJ81 seems to be a restriction site polymorph of λJ19, showing five *Hind*III bands of 4.3, 2.3, 1.45, 0.6 and 0.52 kb (Fig. 4). The 2.3-kb band is readily detected in the genomic blot by a pLAV13 probe, although the 4.3-kb fragment is not. The finding that nick-translated λJ19 DNA hybridizes to all five *Hind*III bands of λJ81 in stringent hybridization and washing conditions indicates that λJ81 is a *Hind*III variant and not a recombinant virus. Also, other mapped restriction sites in λJ81 are identical to those of λJ19 (not shown). Thus, the *Hind*III restriction pattern in the Southern blot can be explained by variation within the single isolate of LAV used to infect the T cells.

HTLV-I[15] and HTLV-II[16] constitute a pair of C-type transforming retroviruses with a tropism for the T-cell subset, OKT4. Both genomes (comprising one LTR) are ~8.3 kb long[17,18], have an X region and show extensive sequence homology. They hybridize between themselves in reasonably stringent conditions (40% formamide, 5 × SSC) and the X regions hybridize even at 60% formamide[19]. Thus, a conserved X region is a hallmark of

this class of virus. We have compared cloned LAV DNA and cloned HTLV-II DNA (pMO)[29] by blot-hybridization and find no cross-hybridization in low stringency conditions of hybridization and washing (T_m = 55 °C), even after 2 days' exposure at −70 °C using intensifying screens (data not shown).

The human T-lymphotropic retroviruses HTLV-III[4] and ARV[12], recently isolated from patients with AIDS or pre-AIDS, have similar morphological, biochemical and immunological properties to LAV, which suggests that they probably represent different isolates of the LAV prototype. DNA hybridization between HTLV-III and HTLV-I and -II has been reported, most noticeably at the *gag-pol* junction and less so in the characteristic X region of HTLV-I and -II[21]. As mentioned above, we could detect no such hybridization and conclude that the reported homology must have been due to either (1) the use of an uncloned cDNA as hybridization probe, (2) the fact that the isolates in question differ substantially from those we have cloned, or (3) the possibility that HTLV-III and a HTLV-I/II-like virus were co-infecting the cells. The last possibility may also apply to the preliminary report of cross-hybridization between a LAV-like virus and a cloned HTLV-II DNA probe[7]. Thus, we find no molecular evidence of a relationship between LAV and HTLV. Furthermore, the LAV genome is ~9 kb long, compared with 8.3 kb for the HTLV viruses[17,18]. Despite their comparable genome sizes, LAV does not cross-hybridize with Visna virus[22] (~9 kb) (data not shown) or with several human endogenous viral genomes (ref. 23 and M. Martin, personal communication) in non-stringent conditions (T_m = 55 °C). These data and morphological and immunological dissimilarities[1,2] between LAV and the HTLV-I/-II pair all point to LAV being a novel class of human retrovirus.

In conclusion, we have molecularly cloned the complete genome of LAV from freshly infected activated T cells of a healthy donor. It has been shown that the tropism of certain retroviruses resides in the LTR[24,25] and that sequence differences and insertions/deletions are present in the LTRs of leukaemogenic and non-leukaemogenic retroviruses. It is thus possible that LAV and LAV-like viruses passaged through B- and T-transformed cell lines[4,12,13] might have undergone some attenuation. Although the cDNA clones were made from a LAV-producing B-cell line, the genomic clones were isolated from LAV-infected normal T cells. Thus, the clones represent LAV genomes that have not been selected or adapted to a particular cell line. However, the LAV genome is shown to be polymorphic even within a single isolate and independent isolates will probably differ widely.

The availability of cloned LAV DNA should facilitate the understanding of the molecular mechanism of viral replication, and the tropism of the virus. The DNA sequence of LAV opens up the possibility of expressing the viral *gag* and *env* gene products and of studying the molecular basis of LAV antigenicity.

We thank Drs D. Dormont and J. Weissenbach for their interest in this work, Denise Guetard, Sophie Chamaret and Jacqueline Gruest for cells, Dr R. C. Gallo for the HTLV-II probe (pMO), Dr M. Brahic for a cloned Visna probe (A109) and Ana Cova for typing the manuscript. This work was supported by grants from the CNRS, the Association pour la Recherche contre le Cancer, the Fondation pour la Recherche Médicale and Institut Pasteur.

Received 30 September; accepted 12 November 1984.

1. Barré-Sinoussi, F. et al. *Science* 220, 868–871 (1983).
2. Pinching, A. J. *Clin. exp. Immun.* 56, 1–13 (1984).
3. Montagnier, L. et al. in *Human T-cell Leukemia Viruses* (eds Gallo, R. C., Essex, M. & Gross, L.) 363–379 (Cold Spring Harbor, New York, 1984).
4. Vilmer, E. et al. *Lancet* i, 753–757 (1984).
5. Feorino, P. M. et al. *Science* 225, 69–72 (1984).
6. Klatzmann, D. et al. *Science* 225, 59–63 (1984).
7. Kalyanaraman, V. S. et al. *Science* 225, 321–323 (1984).
8. Popovic, M., Sarngadharan, M. G., Read, E. & Gallo, R. C. *Science* 224, 497–500 (1984).
9. Gallo, R. C. et al. *Science* 224, 500–503 (1984).
10. Schüpbach, J. et al. *Science* 224, 503–505 (1984).
11. Sarngadharan, M. G., Popovic, M., Bruch, L., Schüpbach, J. & Gallo, R. C. *Science* 224, 506–508 (1984).
12. Levy, J. A. et al. *Science* 225, 840–842 (1984).
13. Montagnier, L. et al. *Science* 225, 63–66 (1984).
14. [illegible] 249–256 [illegible]
15. Gallo, R. C. et al. *Proc. natn. Acad. Sci. U.S.A.* 79, 5680–5683 (1982).
16. Kalyanaraman, V. S. et al. *Science* 218, 571–573 (1982).
17. Seiki, M., Hattori, S., Hirayama, Y. & Yoshida, M. *Proc. natn. Acad. Sci. U.S.A.* 80, 3618–3622 (1983).
18. Chen, I. S. Y., McLaughlin, J., Gasson, J. C., Clark, S. C. & Golde, D. W. *Nature* 305, 502–505 (1983).
19. Shaw, G. M. et al. *Proc. natn. Acad. Sci. U.S.A.* 81, (in press) (1984).
20. Gelmann, E. P., Franchini, G., Manzari, V., Wong-Staal, F. & Gallo, R. C. *Proc. natn. Acad. Sci. U.S.A.* 81, 993–997 (1984).
21. Arya, S. K. et al. *Science* 225, 927–930 (1984).
22. Harris, J. D. et al. *Virology* 113, 573–583 (1981).
23. Steele, P. E., Rabson, A. B., Bryan, T. & Martin, M. A. *Science* 225, 943–947 (1984).
24. Lenz, J. et al. *Nature* 308, 467–470 (1984).
25. Chen, I. S. Y., McLaughlin, J. & Golde, D. W. *Nature* 309, 276–279 (1984).
26. Gubler, U. & Hoffman, B. J. *Gene* 25, 263–269 (1983).
27. Grunstein, M. & Hogness, D. *Proc. natn. Acad. Sci. U.S.A.* 72, 3961–3965 (1975).
28. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* 74, 5463–5467 (1977).
29. Ish-Horowicz, D. & Burke, J. F. *Nucleic Acids Res.* 9, 2989–2998 (1981).
30. Benton, W. D. & Davis, R. W. *Science* 196, 180–182 (1977).

Molecular cloning of AIDS-associated retrovirus

Paul A. Luciw, Steven J. Potter, Kathelyn Steimer & Dino Dina

Chiron Research Laboratories, Chiron Corporation, Emeryville, California 94608, USA

Jay A. Levy

Cancer Research Institute, Department of Medicine, University of California, San Francisco, California 94143, USA

Retroviruses cause a wide variety of diseases in avian and mammalian species. Human acquired immune deficiency syndrome (AIDS) leads to collapse of the immune system and death by a wide variety of opportunistic infections; unusual forms of cancer are associated with this syndrome. Retroviruses have been recovered from tissues of AIDS patients and from patients with related conditions. These similar newly-isolated viruses are lymphadenopathy-associated virus (LAV)[1], human T-cell lymphotropic virus (HTLV-III)[2,3] and AIDS-associated retrovirus (ARV-2)[4]. We have identified a RNA genome of ~9 kilobases (kb) in virions purified from the culture medium of a human T-cell tumour line infected with ARV-2. A cDNA probe made from viral RNA detected circular DNA molecules and proviral forms in infected cells. We prepared a library of infected cell DNA. Recombinant phage included those with a 9.5-kb proviral DNA and viral DNA permuted with respect to the single *Eco*RI site. Comparison of three ARV isolates from different AIDS patients revealed polymorphism of restriction endonuclease sites.

HUT-78 cells, originating from a human T-cell lymphoid tumour[5], were used to propagate the ARV-2 strain of virus[4]. To characterize the viral genome, RNA was extracted from purified virions and electrophoresed on agarose gels containing methyl mercury hydroxide[6]. A distinct ~9-kb RNA species was observed (Fig. 1) with smaller heterogeneous RNA and some ribosomal RNA species. The 9-kb RNA species was used as a template with random primers in a reverse transcriptase reaction to produce a virus-specific cDNA probe[7]. RNA of virus obtained from cells infected with ARV-2 or with two additional isolates, ARV-3 and ARV-4, showed distinct bands at 9 kb that hybridized with the cDNA probe (Fig. 1).

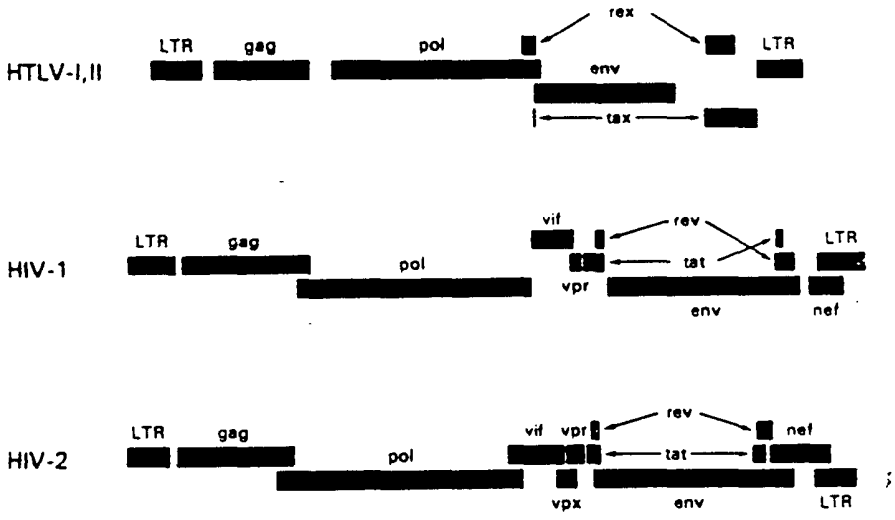
With this cDNA probe, we examined the structure of viral DNA in infected cells by digestion with restriction enzymes, electrophoresis in agarose gels and Southern blotting. No specific bands were detected in several digests of DNA from uninfected cells (Fig. 2*a*, lanes *C*, *E*), whereas bands were seen in infected cells (Fig. 2*a*, lane *A*). Undigested DNA from infected cells contained a species at 5.5 kb, a faint species at 6 kb

HIV/HTLV gene nomenclature

SIR—The complexities of the genomes of human retroviruses (the human T-cell leukaemia viruses, HTLV-I and HTLV-II, and the AIDS-causing human immunodeficiency viruses, HIV-1 and HIV-2) are being unravelled at a rapid pace which is likely to continue and expand. In addition to containing a large ensemble of positive and negative regulatory genes that orchestrate virus expression, these viruses are also remarkable in that they seem to have converged onto parallel regulatory pathways. Two of the regulatory genes of the immunodeficiency viruses are analogous to the two regulatory genes of the leukaemia viruses, although their detailed mechanisms of action may be quite different. Deciphering the modes of action of the regulatory genes of these viruses is crucial to the understanding of their pathogenesis as well as to development of therapeutic agents. Because of the tremendous activity in this field, more than one name has sometimes been given to a single gene and the same name may also apply to more than one gene. In the interest of the many new investigators entering the field for the first time, we feel it is important that we reach a standard nomenclature for all known genes of HIV and HTLV. We propose the scheme outlined in the table.

ROBERT GALLO
FLOSSIE WONG-STAAL
National Cancer Institute, NIH, Bethesda, Maryland 20892, USA
LUC MONTAGNIER
Department of Virology, Institut Pasteur, 75724 Paris Cedex 15, France
WILLIAM A. HASELTINE
Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA
MITSUAKI YOSHIDA
Department of Viral Oncology, Cancer Institute, Tokyo 170, Japan

Proposed name (and derivation)	Previous names	Molecular mass ($\times 10^{-3}$)	Known function
HTLV-I and HTLV-II genes:			
tax$_1$ (transactivator)	*x-lor, p40x, tat*$_1$	41, 41, 42	Transactivator of all viral proteins
tax$_2$	*tat*$_2$, *TA*	38	
rex$_1$(regulator of expression	*pp27x, tel*	27	Regulates expression of virion proteins
rex$_2$ virion proteins)		25	
HIV genes:			
tat (transactivator)	*tat-3, TA*	14	Transactivator of all viral proteins
rev (regulator of expression of virion proteins)	*art, trs*	19, 20	Regulates expression of virion proteins
vif (virion infectivity factor)	*sor, A, P', Q*	23	Determines virus infectivity
vpr (R)	*R*	?	Unknown
nef (negative factor)	*3'orf, B, E', F*	27	Reduces virus expression, GTP-binding
vpx (X) (only in HIV-2 and SIV)	*X*	16, 14	Unknown



Vpr and *vpx* are temporary names and may be changed when more information about their functions is available. Subscripts 1 and 2 would be used to distinguish genes of HIV-1 and HIV-2 (for example, *rev*$_1$ and *rev*$_2$). It is expected that genes of the simian viruses (STLV-I, SIV) would follow similar nomenclature with the subscripts STLV or SIV as appropriate.

Estimating the incubation period for AIDS patients

SIR—The nonparametric analyses of the data on transfusion-related AIDS considered by Medley *et al.*[1] indicate problems of identifiability. With data obtained by retrospective determination of the time of infection for diagnosed AIDS cases, it is only possible to estimate the early part of the incubation distribution up to a constant of proportionality. The same applies to the total number of infections by blood transfusion before any given time. The transfusion data themselves are unable to discriminate between high infection rates coupled with long incubation times on the

As do Medley *et al.*[1], we postulate a function $h(x)$ which specifies the increase over time of the number of HIV-infected individuals who eventually develop AIDS, and a probability density function $f(s)$ for the incubation time of those individuals. The corresponding likelihood function can be maximized jointly with respect to h and f. As the likelihood depends only on the product of h and f, it is not possible to estimate either of these fuctions completely; they may be individually estimated only up to constants of proportionality c and c^{-1}, respectively. nosed within t years of infection, $F(t) = \int_0^t f(u)du$, are given in the figure for the three age groups considered by Medley *et al.*. In this figure we show the estimates of $F(t)$ so that for each group, $c = F(7.5)$. For the children, the levelling of the estimate of $F(t)$ by about 3.5 years suggests that the whole of the distribution of incubation times has been seen; it may then be reasonable to suppose that $c = 1$ but, as also noted by Medley *et al.*, a second wave of incubation times that exceed 7.5 years is not excluded by these data. For the other two age groups, there is nothing in the transfusion data themselves to suggest a value for c. As a consequence, it is impos-

nature

16 April 1987
Volume 326
Issue no. 6114

Micrographs of an icosahedral 'flower' obtained by solidification of an Al–Li–Cu alloy were generated by Professor Guinier on an image processor starting from a scanning electron micrograph and using pseudo-colours. See News and Views p.640.

THIS WEEK

■ New AIDS virus
. . . page 662

A second AIDS virus, HIV-2, is shown by sequencing to be only distantly related to HIV-1.

■ AIDS: activation
. . . page 711

HIV-1 activation is triggered by a protein of the T-cell immune response.

■ Feedback loop controls climate
. . . page 655

A chemical released by marine plankton indirectly causes cloud formation, creating a feedback loop in which plankton and climate keep each other in balance.

■ Release of tamarins
. . . page 642

Jesse Cohen

Successful reintroduction of golden lion tamarins into the wild requires better pre-release training.

■ Rats should not shave
. . . page 694

Trimming a rat's whiskers from the day of birth causes abnormalities in the cortex of the adult brain.

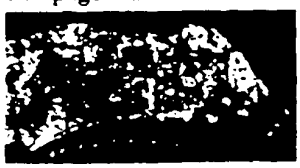
■ Superconductors
. . . page 669

New superconducting materials show unexpected switching in their magnetic properties.

. . . page 638

How will superconductors be used? And the advances being made in Japan.

■ A new fossil monkey
. . . page 692



A new find shows a remarkable evolutionary and behavioural conservatism in the owl monkey lineage.

■ New space structures
. . . page 675

Are unexpected long-period fluctuations in extra-galactic radio-frequency signals hitherto undetected structures in the interstellar medium?

■ Supernova 1987A
. . . page 638

Computer models may identify the star which exploded to form the supernova star.

■ Atrial natriuretic peptide in the kidney
. . . pages 644 and 697

Peptide inhibition of angiotensin II blocks reabsorption of sodium and water in the renal tubule.

OPINION

NEWS

NEWS & VIEWS

SCIENTIFIC CORRESPONDENCE

COMMENTARY

BOOK REVIEWS

Contents continued ▶

Nature® (ISSN 0028-0836) is published weekly on Thursday, except the last week in December, by Macmillan Magazines Ltd and includes the Annual Index (mailed in February). Annual subscription for USA and Canada US $250. USA and Canadian orders to: *Nature*, Subscription Dept, PO Box 1501, Neptune, NJ 07753, USA. Other orders to *Nature*, Brunel Road, Basingstoke, Hants RG21 2XS, UK. Second class postage paid at New York, NY 10012 and additional mailing offices. Authorization to photocopy material for internal or personal use, or internal or personal use of specific clients, is granted by *Nature* to libraries and others registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided the base fee of $1 per copy plus $0.10 per page is paid directly to CCC, 21 Congress Street, Salem, Massachusetts 01970, USA. Identification code for *Nature*: 0028-0836/87 $1.00 + $0.10. US Postmaster send address changes to *Nature*, 65 Bleecker Street, New York, NY 10012. Published in Japan by Macmillan Shuppan K.K., Eikow Building, 1-10-9 Hongo, Bunkyo-ku, Tokyo, Japan.

Genome organization and transactivation of the human immunodeficiency virus type 2

Mireille Guyader, Michael Emerman, Pierre Sonigo, François Clavel*, Luc Montagnier & Marc Alizon†

Unité d'Oncologie Virale (CNRS UA 1157), and Unité de Recombinaison et Expression Génétique (INSERM U163, CNRS UA271), Laboratoire de Biologie Moléculaire et Immunologie des Rétrovirus, Institut Pasteur, 25 rue du Dr Roux 75724 Paris Cedex 15, France

Analysis of the nucleotide sequence of the human retrovirus associated with AIDS in West Africa, HIV-2, shows that it is evolutionarily distant from the previously characterized HIV-1. We suggest that these viruses existed long before the current AIDS epidemics. Their biological properties are conserved in spite of limited sequence homology; this may help the determination of the structure-function relationships of the different viral elements.

THE acquired immune deficiency syndrome (AIDS) has now spread worldwide and appears to be an acute public health problem in Africa in particular[1-5]. A retrovirus designated human immunodeficiency virus (HIV), but previously known as LAV, HTLV-III or ARV, was shown to cause AIDS in the different areas afflicted by the epidemics[6-8]. Indeed, isolates from North America, Western Europe and Central Africa have the same biological properties, and antigenically cross-reactive proteins with the same relative molecular mass[9-11]. Only studies at the molecular level have revealed some differences in the nucleotide sequence of North-American and African isolates[12,13]. This sequence variation is also present, though to a lesser extent, among different isolates from the USA[14-18].

The western part of Africa seemed relatively spared by AIDS[3]. Recently, however, several typical cases were found in a survey of patients from Guinea Bissau and other countries of West Africa[19-21]. Unexpectedly, most of these patients did not have detectable titres of antibodies against HIV. But they were found to be infected by a retrovirus related to HIV by its ultrastructural and biological properties, such as cytopathogenicity and tropism for cells carrying the CD4(T4) antigen[19]. Antibodies raised against HIV could immunoprecipitate the *gag* and *pol* products of these isolates, which have molecular masses that are similar but not identical to these antigens of HIV; in contrast, the *env* products could not be immunoprecipitated, whereas previous HIV isolates showed wide cross-antigenicity of the envelope glycoprotein. Furthermore, the genome of this new retrovirus cross-hybridized only poorly in very low stringency conditions with HIV DNA probes[19,22]. We have therefore designated this West African AIDS virus as HIV type 2 (HIV-1 referring to the AIDS retrovirus previously identified in Central Africa, North America and Europe). More than 20 isolates have so far been made from patients with AIDS and related conditions, mainly originating from west Africa[20,21], but also in some Europeans (L.M., unpublished), and epidemiological studies in progress indicate a seroprevalence of 1-2% in some populations of West Africa (F. Brun-Vézinet, personal communication).

HIV-2 appears to be closely related to the simian immunodeficiency viruses (SIV) a group of cytopathic retroviruses whose prototype, STLV-3_{mac}, was identified in captive rhesus monkeys (*Macaca mulatta*) with an AIDS-like disease[23], and was later found to infect other primate species, either wild or in captivity[24-26]. Genetic comparisons of SIV, HIV-1 and HIV-2 may help to elucidate the phylogeny of these viruses and the origins of the recent AIDS epidemics. As these retroviruses share most of their biological properties, the identification of conserved sequences is important to localize the functional domains of the viral proteins and regulating elements, and design new diagnostic and therapeutic tools. We present here the complete nucleotide sequence of HIV-2, the comparison of its proteins with those of HIV-1, and preliminary studies on the regulation of HIV-2 expression.

Nucleotide sequence and LTR analysis

The sequence presented in Fig. 2 is derived from two λ clones corresponding to integrated proviral DNA from the ROD isolate of HIV-2 (ref. 22), obtained in 1985 from an AIDS patient from Cape Verde Islands (offshore Senegal, refs 19, 20). The genome of HIV-2 is 9,671 nucleotides long (in its RNA form), whereas HIV-1 isolates are about 9,200 nucleotides long. This difference is partly explained by the respective sizes of the long terminal repeats (LTRs, see below).

The genetic organization of HIV-2 (shown in Fig. 1) is analogous to that of HIV-1, that is:

5'LTR-*gag*-*pol*-central region-*env*-orf F-3'LTR.

The 'central region', also identified in the ovine lentivirus visna[27], contains five major open reading frames (ORFs), four being clearly related to the ORFs of HIV-1 that encode the Q (or *sor*), R, *tat* and *art* (or *trs*) genes of HIV-1 (refs 15-18, 27-31). The fifth, which we designate ORF X, has no obvious counterpart in HIV-1. Alignments of the nucleotide sequences of HIV-1 and 2 show their distant homology (from ~60% for the more conserved *gag* and *pol* genes, to 30-40% for the other viral genes and LTRs). To allow these alignments to be made many insertions and deletions must be introduced into the sequences. We do not find that these insertions are the small duplications that would be characteristic of the recent divergence of retroviral sequences, as was noted among isolates of HIV-1 (ref. 12).

The limits of the LTRs and of their internal U3, R and U5 elements, determined by sequence analysis and some complementary experiments, are shown in Fig. 2. Classically bounding the retroviral LTRs are short inverted repeats (5' CTG-CAG 3') located after a polypurine tract for the 3'LTR, and before a sequence complementary to the 3' end of a transfer RNA that is used as primer by the reverse transcriptase (here, as in HIV-1 and visna virus, a lysine tRNA, refs 15, 27) for the 5' LTR. The R-U5 junction, corresponding to the 3' end of the polyadenylated viral RNA, was previously localized by sequencing oligo(dT)-primed complementary DNA (cDNA) derived from the HIV-2_{ROD} genome[22]. The length of U5+R, and hence the position of the U3-R junction corresponding to the 5' cap site of the viral RNA were deduced from the size of a HIV-2 cDNA synthesized using the endogenous reverse transcriptase activity and the endogenous $tRNA^{lys}$ primer (see Fig. 3). This

* Present address: Laboratory of Molecular Microbiology, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland, USA
† To whom correspondence should be addressed.

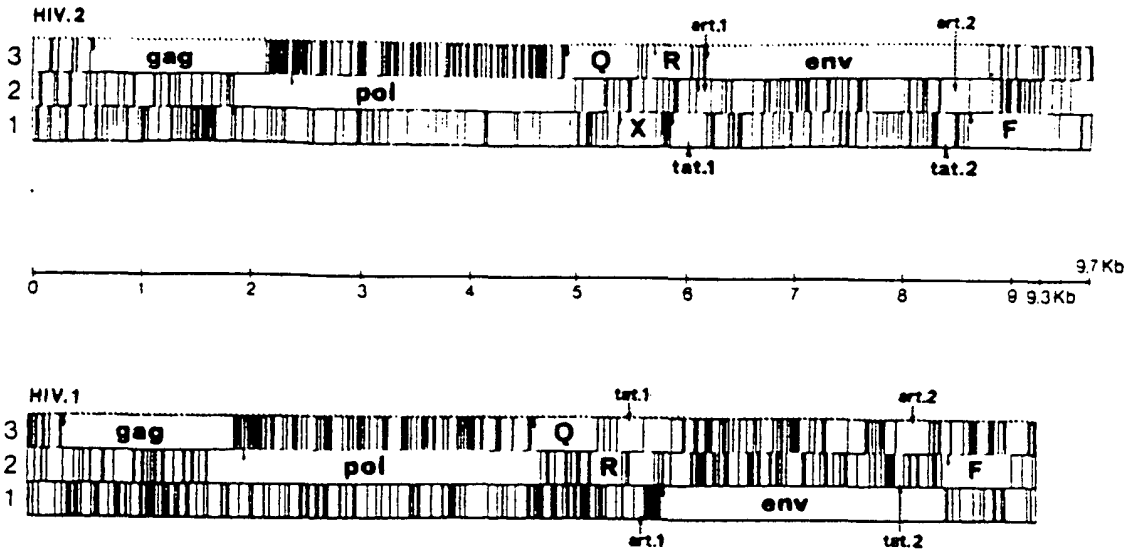


Fig. 1 Organization of the HIV-2 and HIV-1 genome (BRU isolate, ref. 15). Vertical bars represent the stop codons in the 3 reading frames. Arrows indicate the initiator AUG codons in viral genes or potential genes. *Tat* 1 and 2, *art* 1 and 2 are the open reading frames containing the coding exons of the *tat* and *art* genes.

strong-stop cDNA' is 302 +/−1 nucleotides long (181 nucleotides in HIV-1, ref. 15). Thus, the U5 element is 125 bp long, U3 is 556 bp and R 173 bp (respectively 82, 456 and 97 bp in HIV-1). All the elements of the HIV-2 LTRs are larger than in HIV-1, and alignment by computer programs shows large insertions and very distant overall homology for the aligned regions[22]. However, the three Sp1 binding sites identified in HIV-1 (ref. 32), are also present in HIV-2 from nucleotide 9,419 to 9,448 with 17 out of 29 nucleotides homologous to this region of HIV-1. The core enhancers identified in HIV-1 (ref. 33) are present in HIV-2 from nucleotide 9,389 to 9,416: the first is 50% homologous and the second 100% homologous to that in HIV-1 (Fig. 2).

The analysis of the virus-specific poly(A)$^+$ RNA (not shown) from a cell line infected with and continuously producing HIV-2 revealed a pattern of transcription reminiscent of that observed in HIV-1-infected cells: RNA of over 9 kilobases (kb), corresponding to a full-length transcript, and three types of spliced messenger RNA of 5, 4.5 and 2 kb, also observed in HIV-1 (refs 18, 34).

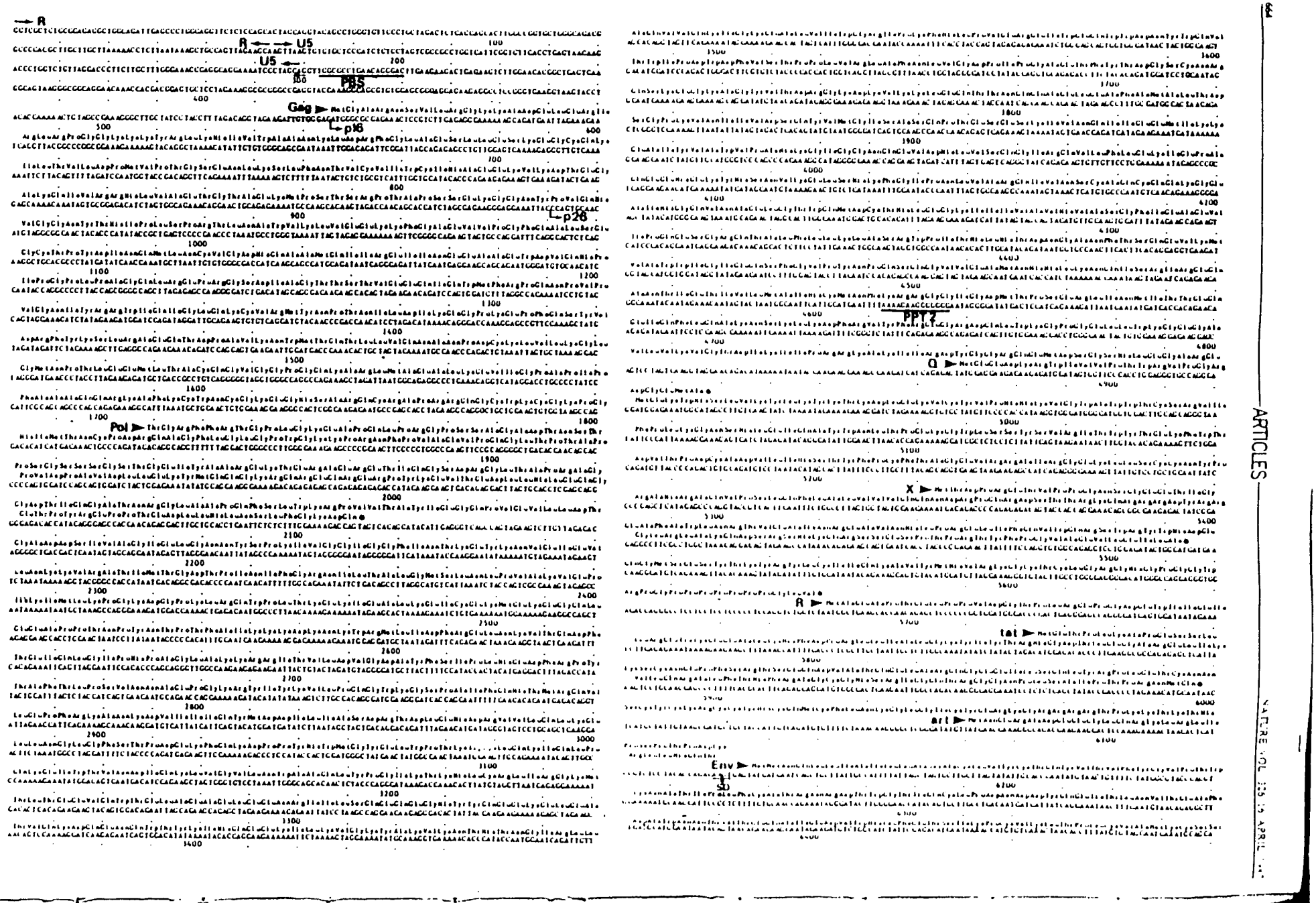
The *gag* and *pol* proteins and HIV phylogeny

The *gag* precursor of HIV-2 has a calculated relative molecular mass of 57,100 (M_r 57.1K), consistent with the p55 antigen[20] seen by immunoprecipitation with patient sera, and is probably processed, by analogy with HIV-1, into the proteins designated p16, p26 and p12 (refs 19, 20). By analogy with the p18^{gag} of HIV-1, p16 would be at the amino terminus of *gag* and precede p26, whose amino terminus has been sequenced (H. Marquardt, personal communication) and starts with the proline residue at position 951. The carboxy-terminal part of the *gag* precursor encodes a p12 that contains the cysteine-rich consensus of the retroviral nucleic-acid-binding proteins also found twice in the p13^{gag} of HIV-1 (ref. 15). The HIV-2 *pol* ORF could encode the p64 and p36 antigens of HIV-2 (ref. 20) which by analogy correspond to the p68 and p34 (reverse transcriptase and endonuclease, respectively[35]) of HIV-1.

The *gag* and *pol* proteins of HIV-1 and 2 were expected to share large conserved domains, as these HIV-2 proteins can be precipitated by antibodies in sera from patients infected with HIV-1. However, we found that only 58% and 59.4% of the amino acids of *gag* and *pol* respectively are identical to the corresponding HIV-1 products (Table 1*a*), whereas the more distant isolates of HIV-1 (Zairian and US) show 90 to 95% amino-acid identity in these proteins (Table 1*b* and ref. 12). Several insertions and deletions have to be introduced in the alignments (data not shown), whereas they are rare in the comparisons of *gag* and *pol* genes between HIV-1 isolates. The *gag* and *pol* proteins of HIV-2 are no closer to those of the Zairian isolates than to the prototype HIV-1 (BRU isolate) isolated in 1983 from a French patient[6] probably infected in the USA. Overall, the difference in *gag* and *pol* between HIV-1 and HIV-2 is of the same order as that observed among the group of the human T-cell leukaemia viruses (HTLV-I and II) and bovine leukaemia virus (BLV). However, this latter group displays a higher conservation in the envelope, 70% amino-acid identity between HTLV-I and HTLV-II, versus about 42% between HIV-1 and HIV-2 (see below). Alignments of different retroviral *pol* proteins (Table 1*b*) confirm that the HIVs form a subgroup that is more related to the lentiviruses visna and equine infectious anaemia virus (EIAV) than to any other human or animal retrovirus.

Homologous domains in *env*

The envelope glycoproteins of retroviruses are translated from a subgenomic viral mRNA (here probably the transcript of 4.5 kb). Addition of sugar residues (*N*-linked glycosylation) gives rise to a high-M_r precursor which is processed by proteolytic cleavage. The length of the leader sequence of the HIV-2 glycoprotein cannot be precisely determined by alignment with that of HIV-1 (experimentally found to be 32 amino acids long[36]) because of a lack of sequence homology (Fig. 4). But the amino terminus of *env* contains a relatively hydrophobic stretch in the calculated hydropathy plot (not shown) that is probably the signal peptide. The potential cleavage site between the external envelope glycoprotein (120K) and the transmembrane protein (previously thought to be the 36K antigen[19], and now putatively identified as a 40K antigen[20]) is found at amino acid 505 (Fig. 4) immediately after the Lys-Glu-Lys-Arg sequence. This cleavage site aligns partly to one (Lys-Ala-Lys-Arg) of the two potential cleavage sites found in HIV-1 (the other being located after the Arg-Glu-Lys-Arg stretch). The calculated M_r of the extracellular glycoprotein (EGP) and of the transmembrane protein (TMP) of HIV-2 would be 57K and 41.7K respectively; the discrepancy


R
U5
PBS
Gag
p16
p28
Pol
PPT
X
Q
tat
art
Env
SD

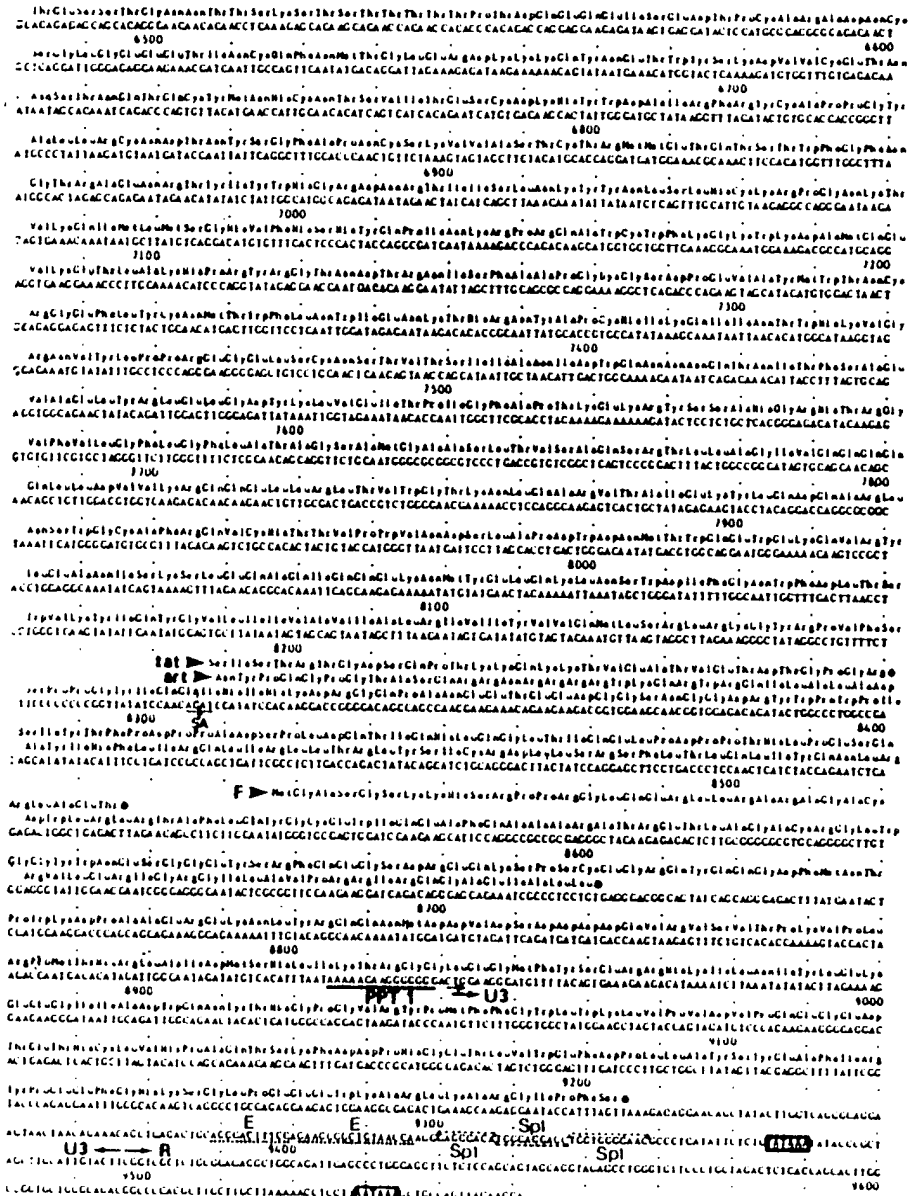


Fig. 2 Complete nucleotide sequence of the HIV-2_{ROD} provirus genome. The sequence of the 9,671 nucleotide coding strand from the cap site to the polyadenylation site of the RNA is shown together with the deduced amino-acid sequence of the viral proteins. The 3-bp inverted repeats bounding the LTR are underlined with arrows. The primer binding site (PBS) complementary to the 3' end of $tRNA^{lys}$, and polypurine tract (PPT) are underlined. The repeat of the PPT in the *pol* gene is also underlined. The limits of the U3, R and U5 elements of the LTR are indicated by arrows. The promoter (TATAAA), the three potential Sp1 factor binding sites (indicated by dotted lines) and the two core enhancer sequences (E, indicated by dashed lines) of U3 are shown; in R the polyadenylation signal AATAAA is underlined. The viral genes and potential genes are translated from the first AUG of the open reading frame (ORF), except *pol* which is translated from the beginning of the ORF. The end of translation is indicated by a filled circle. The amino terminus of the major core protein, the $p26^{gag}$ was determined by protein sequencing. SD at position 6,140 and SA at position 8,307 indicate the probable splice donor and acceptor sites of the intron separating the two coding exons of the *tat* and *art* genes.

Methods. The sequence was determined by the M13 shotgun cloning and dideoxynucleotide chain terminator method[57,58], as described[27] starting from the inserts of the λ phages ROD 27 and ROD 35 containing integrated proviral DNA[22]. ROD 27 corresponds to the 5' part of the genome, to the *Eco*RI site at position 2,658, whereas ROD 35 corresponds to the part of the genome 3' to this site. An in-frame TAG stop codon was found in ROD 35 *env* gene. The sequencing of the corresponding region in an oligo(dT)-primed HIV-2 cDNA (clone E2, described in ref. 22) revealed that it was due to a C-to-T mutation at position 8,304.

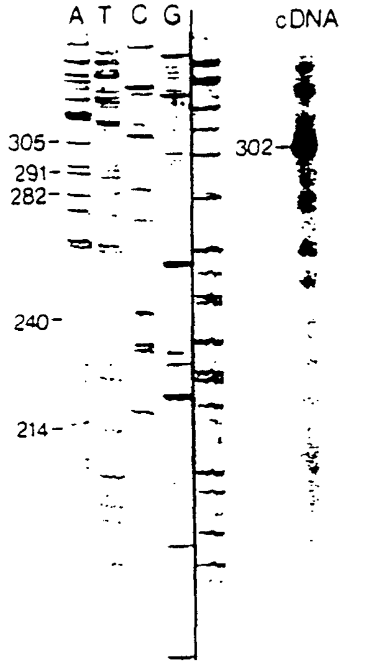


Fig. 3 HIV-2 strong-stop cDNA corresponding to the length of the R–U5 elements of the HIV-2 LTR. The methods were previously described[15]. Briefly, virions were purified by ultracentrifugation and an endogenous cDNA reaction performed with radiolabelled nucleotides after mild disruption of viral envelope with Triton X-100. The tRNA[lys] primer, complementary to the PBS site flanking the U5 element at the 5' end of the genome, was then degraded by alkaline hydrolysis and the cDNA run on a denaturing 6% acrylamide-urea gel together with a sequence reaction for accurate estimation of the size of the products.

with the apparent M_r of the EGP is explained by glycosylation (30 sites in HIV-2, about half of which are conserved with respect to HIV-1).

Figure 4 shows an alignment of the envelopes of the two HIVs. The proteins are overall very distantly related (41.7% identity in the entire envelope, 39.4% in the EGP, 44.8% in the TMP) compared to divergent isolates of HIV-1 (about 75–80% identity in the whole envelope, ref. 12). Many large insertions have to be introduced, particularly in alignment of the EGPs where only short, widely separated domains are conserved between HIV-1 and 2. These domains are clustered into the conserved regions of the EGP of HIV-1 (identified by comparison of different isolates[12-14]), and generally coincide with cysteine residues. Among the HIV-1 isolates, all the cysteine residues could be aligned in spite of the generally large genetic variation, especially in gp110. Almost all (22/23) of the cysteine residues of HIV-1 can also be aligned with HIV-2, but the latter contains seven additional cysteine residues, often in the regions representing insertions relative to HIV-1. Thus, the folding of the HIV-2 EGP could be different from that of HIV-1, and some regions, therefore, might be exposed in a different manner.

Other viral proteins

The HIV-1 genome contains several other genes encoding proteins of small M_r (10 to 27K), two of which (*tat* and *art/trs*) have an identified function: the positive regulation of viral expression[30-33]. No role has yet been identified for the p23 encoded by ORF Q (or *sor*)[37,38], nor for the p27 encoded by ORF F (or 3' ORF)[39]. We also observed in the region between the *pol* and *env* genes of HIV-1 (central region) another potential gene, which we designated R (ref. 12). All these elements are found in HIV-2, but the corresponding proteins are only distantly homologous (see Table 1*a*). In the F protein, most of the difference between HIV-1 and 2 is due to a large insertion in the amino terminus of HIV-2. The second half of the protein, encoded by the U3 element of the LTR, shows better conservation (data not shown).

Based upon sequence homologies with HIV-1, the *tat* and *art* genes of HIV-2 are probably organized as split genes transcribed into ~2 kb mRNA made of three exons[18,28-31]: the 5' leader, a first coding exon located in the central region and probably ending at a possible splice donor found at position 6,1– (CAAGT, Fig. 2), and a last exon probably starting at the splice acceptor at position 8,307 in HIV-2 (CAGATC). The *tat* protein of HIV-2 would be longer than that of HIV-1 (130 versus 86 amino acids), having two large insertions in the amino terminus and in the second coding exon (Fig. 4). The main domain of homology of the *tat* proteins corresponds to a region very rich in cysteine residues whose structure is reminiscent of that of the 'cysteine fingers' of some transcription-regulating elements that interact with nucleic acids, such as the TFIIIA factor[40]. This region is followed by an Arg-Lys-rich stretch that could also interact with DNA or RNA. No significant homology is seen in the second coding exon, which has been shown to be dispensable to the function of the protein[28,29]. The *art*-encoded protein is shorter in HIV-2 than it is in HIV-1 (100 versus 116 amino acids), and most of its length is encoded by the last exon. The most homologous part is located in a stretch of basic residues that may be able to interact with nucleic acids.

Cross-transactivation of HIV-1 and HIV-2

The trans-activator gene (*tat*) has been shown to be indispensable for the replication and cytopathicity of HIV-1 (ref. 41).

Table 1 Quantification of the homologies among retroviral proteins

a

HIV-1	GAG	POL	ENV		F	Central Region			
			EGP	TMP		Q	R	TAT	ART
HIV-2	[illegible] (95.2)	59.4 (96.5)	39.4 (90.6)	44.8 (91.3)	[illegible]	34.5 (87.[illegible])	[illegible]	[illegible]	[illegible]

b

	HIV-2	HIV-1	HTLV-I	VISNA
HIV-1	59.1 (96.4)	—	ND	ND
LAV-Eli	61.6 (96.1)	94 (98.7)	ND	ND
LAV-Mal	59 (95.2)	92 (98.7)	ND	ND
EIAV	43.8 (92)	41.9 (91.5)	ND	46.7 (90.8)
VISNA	43.7 (88.7)	42.2 (94)	ND	—
HTLV-I	34.8 (70.5)	33.3 (70.3)	—	ND
HTLV-II	ND	ND	62.8 (99.5)	ND
BLV	ND	ND	49.5 (93.2)	ND
RSV	35.9 (72.3)	34.5 (76.2)	38.2 (86.4)	ND

The reference protein of each alignment is that listed at the top of the column. Proteins were aligned using the NUCALN program[61] with following parameters: K-tuple 1, window 20, gap penalty 1. Two results are indicated in each case: the amino-acid identity (%) in the aligned domains (that is, excluding the regions of insertion/deletion), and between parentheses the percentage of the length of reference protein that could be aligned. *a*, Homologies between HIV-1 and HIV-2 proteins. For *env*, the calculation was done for the external glycoprotein (EGP, including the signal peptide, whose length is not exactly known in HIV-2), and the transmembrane protein (TMP). *b*, Comparison of the *pol*-encoded proteins of different retroviruses. LAV.Mal and LAV.Eli are Zairian isolates of HIV-1 (ref. 12); EIAV: equine infectious anaemia virus (sequence communicated by Dr S. Aaronson), and visna virus[27] are animal lentiviruses; HTLV-I, HTLV-II, BLV[62-64], related leukaemogenic retroviruses; RSV) Rous sarcoma virus[65]. ND, not determined.

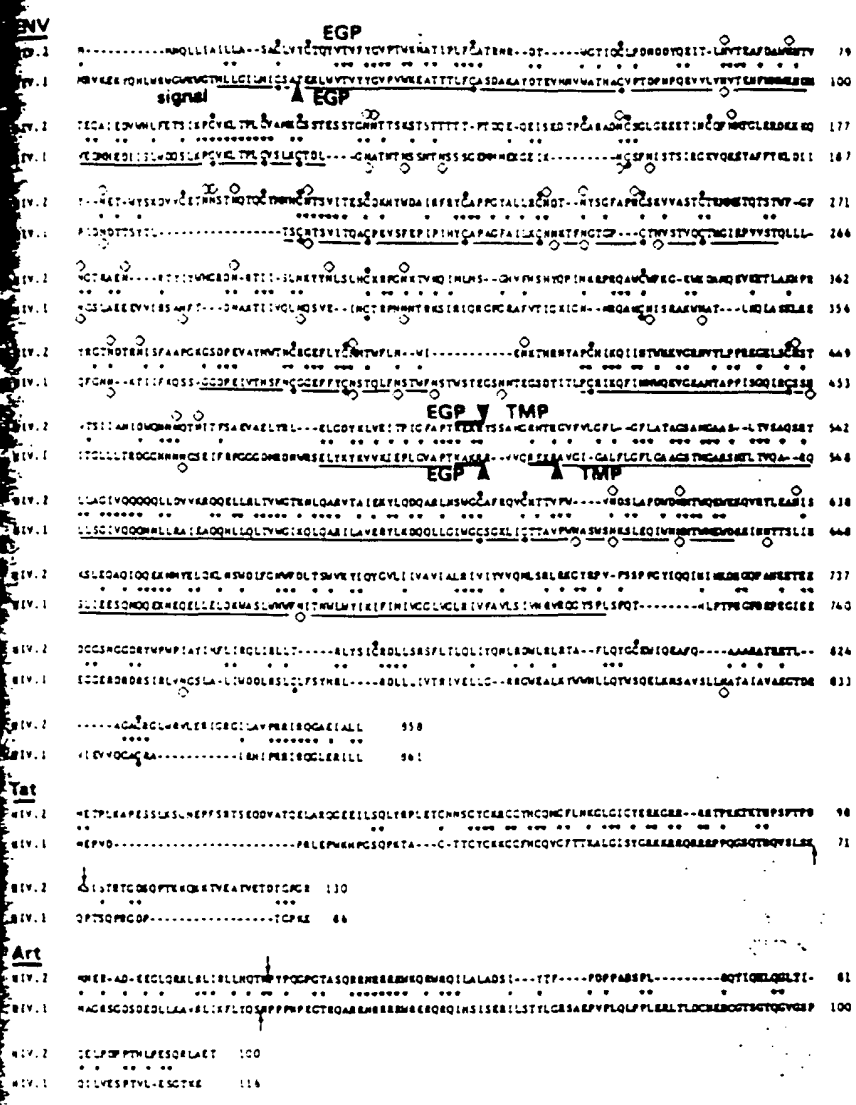


Fig. 4 Alignments of the HIV-1 (BRU isolate, ref. 15) and HIV-2 proteins. Asterisks indicate amino-acid identities. Gaps were introduced to optimize the alignments. In the envelopes, the potential cleavage sites are shown by arrows. EGP, external glycoprotein; TMP, transmembrane protein. ◇, Potential *N*-glycosylation sites; ●, cysteines. The domains of the EGP of HIV-1 that were found to be well-conserved among isolates[12] are underlined. The parts of *tat* and *art* encoded by each of the two exons are separated by an arrow.

To examine whether transactivation (a property also shared with the ovine visna lentivirus but not with the related caprine arthritis and encephalitis virus[42]) exists in HIV-2, we constructed a plasmid, called pHIV2-CAT, containing the bacterial chloramphenicol acetyltransferase (CAT) gene under the control of the U3–R region of HIV-2 (225 bp of U3 and 175 bp of R). To test the transactivation of HIV-2, cells were either infected with HIV-2 or mock-infected, and five days later transfected with either pSVCAT (which contains the CAT gene under control of the SV40 early promoter[43]) or pHIV2-CAT. At the time of transfection, the cells were not producing virus. Nonetheless, we observed a substantial increase in the amount of CAT expression in extracts of HIV-2-infected versus mock-infected cells that had been transfected with pHIV2-CAT (Fig. 5*a*). The expression of the SV40 early promoter was not affected by HIV-2 infection.

To determine whether the *tat* gene of HIV-1 could transactivate the LTR of HIV-2 and vice versa, we cotransfected SW480 cells[44] with subgenomic fragments of HIV-1 or HIV-2 and pHIV2-CAT or a plasmid called pHIV1-CAT, which contains U3–R of HIV-1 (the entire U3 and 70 bp of R) directing transcription of the CAT gene. The plasmid pLET (a gift from Dr S. Wain-Hobson) contains the region of the HIV-1 shown by others to encode the HIV-1 *tat* gene[28,29]. The plasmid pME214, on the other hand, contains HIV-2 sequences between nucleotides 5,786 and 8,571 (Fig. 2), and in particular contains the open reading frames of HIV-2 that share homology with the *tat* gene of HIV-1. In both of these plasmids transcription is driven by the LTR of the respective virus, and the first AUG of the transcript is the first AUG of the putative *tat* gene. It should be noted that both these plasmids also contain the coding potential for the *art* gene.

Although the SV40 early promoter was not affected by either the HIV-1 *tat* nor the HIV-2 *tat* genes, both HIV-1 and HIV-2 LTRs were substantially activated by the HIV-1 *tat* gene (Fig. 5*b*). This is perhaps surprising in view of the difference in size of the R region of HIV-1 (where the transactivator responsive region (TAR) resides[45]) and HIV-2. However 35 of the 58 bases present in the first stem-and-loop secondary structure of the TAR region of HIV-1 are conserved, and an analogous stem-and-loop structure with the first 77 bases of R can be drawn for HIV-2 (ref. 33).

The HIV-2 LTR is transactivated over 100-fold by pME214 (Fig. 5*b*). On the other hand, the HIV-1 LTR is not as well transactivated by this plasmid (~5–20 fold, Fig. 5 and other data not shown). Similar results were obtained after transfection of HeLa and HUT 78 cells (data not shown). These experiments indicate that pME214 encodes a functional *tat* gene. In addition, they indicate that the specificity of the HIV-2 *tat* is somewhat different from that of the HIV-1 *tat*. It will be important to determine whether this observation is isolate-specific.

Origin of human immunodeficiency viruses

We have presented here the complete nucleotide sequence of the retrovirus associated with AIDS in West Africa, HIV-2, and tentatively identified the viral proteins either detected in

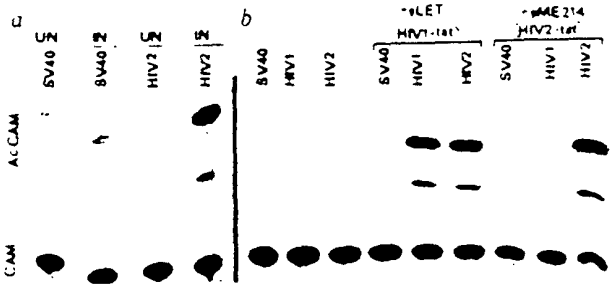


Fig. 5 Transactivation of HIV-2. Chloramphenicol acetyltransferase (CAT) assays were done as described[59]. The unreacted chloramphenicol is marked 'CAM', and the acetylated products are marked 'AcCAM'. All reactions were 1 h with 10% of the cellular extract made 40 h after transfection. The origin of the promoter linked to the CAT gene is indicated above each lane. SV40 indicates the SV40 early promoter, HIV-2 indicates the partial U3 and the entire R sequences of HIV-2 (ROD isolate), and HIV-1 indicates the entire U3 and 70 bp of R of HIV-1 (BRU isolate). *a*, HUT 78 cells were either mock-infected (UN, uninfected) or infected (IN) with HIV-2. Five days post-infection, 3×10^6 cells were transfected with 3 μg of plasmid in 0.5 ml of Tris-saline without divalent cations for 45 min at 37 °C with 250 μg ml^{-1} DEAE-Dextran. *b*, 4×10^5 SW480 cells were cotransfected by the $CaCl_2$ technique[60] with 3 μg of promoter-CAT plasmid and 3 μg of the indicated plasmid. Salmon sperm DNA was added such that each transfection was 20 μg ml^{-1} DNA. This experiment was repeated three times with similar results.

immunoprecipitations with patients' sera, or homologous to proteins previously identified in HIV-1. The two viruses share a similar genomic organization, indicating a common evolutionary origin, but differ significantly in terms of nucleotide and amino-acid sequence: the more-conserved *gag* and *pol* genes respectively display only 56 and 60% nucleotide sequence homology and both less than 60% amino-acid identity. The calculation of the nucleotide sequence homology for the other genes gives even lower values, making HIV-1 and 2 42% homologous overall. This confirms that these two viruses are distinct elements of the HIV family, and cannot be considered as strains of the same virus, according to the recommendations of the international taxonomy committee[46].

It was previously established that HIV-2 is more related to the simian immunodeficiency viruses (SIV) than it is to HIV-1. The *gag*, *pol* and *env* proteins of SIV and HIV-2 are antigenically cross-reactive, whereas their cross-reactivity to HIV-1 is restricted to some *gag* and *pol* antigens. The amino-terminal amino-acid sequence of the major core protein (corresponding to the $p25^{gag}$ of HIV-1 and $p26^{gag}$ of HIV-2) has been determined in one isolate of SIV obtained from macaques with an AIDS-like disease (MnIV, ref. 26). Out of the 23 amino acids sequenced 21 match with the amino terminus of $p26^{gag}$ of HIV-2, whereas 13 (with one deletion) match to the $p25^{gag}$ of HIV-1. Furthermore, whereas HIV-2 can infect, at least transiently, primate species which are evolutionarily more distantly related to humans (at least baboons and macaques), HIV-1 infects only humans and chimpanzees (R. Desrosiers and P. Fultz, personal communications). In fact, it is not possible from current data to know whether SIV can be classified as distinct from HIV-2 or if they only differ as independent isolates of the same virus.

The almost simultaneous emergence of two foci of AIDS in distinct areas of the African continent is unlikely to be due to the recent emergence of two novel human pathogens, for example by simultaneous trans-species infection by animal retrovirus, or by the mutation of pre-existing non-pathogenic human retroviruses. Indeed, HIV-1 and HIV-2 are obviously retroviruses with a common origin, but they are highly divergent, and it is more likely that their time of divergence is earlier than the beginning of the current epidemics. Therefore a common ancestor, with similar properties and pathogenic potential, probably existed a long time ago in a human population, and the emergence of the AIDS epidemics is more likely the result of simultaneous modifications of epidemiological parameters in West and Central Africa, such as uncontrolled urbanization leading to the infection of larger populations.

A question to be addressed is why the HIVs were only recently detected if they existed for a long period. This may be due to the fact that the pathogenicity of an HIV-type retrovirus cannot be revealed until it has spread to a significant portion of the population. First, in areas of Africa with poor medical facilities where other infections, such as malaria, represent primary causes of morbidity, isolated cases of AIDS could have been an undetectable clinical event. Then, the incubation time can vary considerably, and it cannot still be ruled out that a large fraction of individuals infected by a HIV will remain healthy carriers. In Kenya, HIV-1 seropositivity was first reported in a high fraction of subjects at risk of AIDS (female prostitutes) who were apparently healthy; later, the virus diffused to a larger part of the population, and cases of AIDS were observed[47]. A similar situation could explain the apparent lack of pathogenicity of the retrovirus designated HTLV-IV, but indistinguishable from HIV-2 and SIV by the antigenicity of its proteins[3,44,48,49]. The presence of HTLV-IV was identified only in apparently healthy individuals in West Africa, an area where we have observed several typical AIDS cases caused by HIV-2. It is possible that the apparent non-pathogenicity of HTLV-4 is due to a recent epidemic diffusion of HIV-2/HTLV-IV in the West Africa, where AIDS cases still represent a minor fraction of the infected and seropositive individuals, whereas HIV-1 has diffused in major cities of central Africa or the USA some time before.

Implications for vaccines and diagnostics

The risk that HIV-2-infected blood samples may not be detected by standard screens, currently based on the detection of anti-HIV-1 antibodies, makes it important that a way of diagnosing HIV-2 infection is found. As the envelope, and especially its transmembrane part, represents the primary target of the host antibody response to the HIV infection (see ref. 1), antigens from the envelope of HIV-2 will significantly improve the spectrum of the screening tests, allowing the detection of samples infected by HIV-2, and perhaps by other as yet uncharacterized members of the HIV family.

As it shares most of the structural characteristics and biological properties of HIV-1, but displays significant genetic divergence, HIV-2 is a powerful tool in the study of the molecular biology of this group of retroviruses. Among the crucial biological properties common to both HIVs are tropism for CD4-positive cells, and mechanisms of positive regulation of viral expression encoded by viral transactivating factors. We observed that the *tat* of HIV-1 activates the transactivation responsive (TAR) sequences as efficiently in both types of HIV, whereas the *tat* gene of HIV-2 is more efficient on the TAR elements of HIV-2. The *tat* proteins of HIV-1 and 2 have only short homologous sequences, and this will ease the dissection of their function by mutagenesis or using chemically synthesized peptides.

HIV-1 and probably HIV-2 recognize the CD4 surface molecule as a receptor on helper/inducer T lymphocytes and perhaps on other cells expressing the CD4 protein[50-53]. In HIV-1, this interaction is mediated by the external envelope glycoprotein (EGP; ref. 52), and an important problem is which of the domain(s) of this protein are involved in that interaction. Indeed, blocking this step of the virus life cycle, either by antibodies or drugs, could be an efficient means for preventing infection or blocking its spread. As the receptor is a constant cellular protein, we can postulate that the binding domain of the envelope is conserved among the CD4-tropic HIVs. The conserved domains of the EGP of HIV-1 and 2 are not numerous, and therefore it becomes possible to demonstrate their possible role in the virus–receptor interaction using a relatively limited

set of site-directed mutations. Given the absence of antigenic cross-reactivity of the envelopes of the two HIVs, this CD4-binding domain is probably not, or only poorly, immunogenic—perhaps because of masking by glycosylation, poor exposure on the virion surface, or mimicking of 'self' antigens. Nevertheless, its presentation to the immune system out of context of the virion, that is, as a peptide, might induce a neutralizing antibody response that is not attained, or attained with only a low efficiency, with the complete native envelope from virions or expression systems[54-56].

Conclusion

The comparative analysis of HIV-1 and 2 reveals major genetic differences between retroviruses that share many of their biological properties. They both cause AIDS, are cytopathic *in vitro*, have a tropism for CD4-bearing cells and have elements *trans*-activating the expression of viral genes acting at the LTR level. The evolutionary potential of these viruses is therefore striking, and we must ask whether other HIVs can emerge as long as a favourable epidemiological situation is provided. We must take advantage of the precise delineation of the conserved structures to understand their molecular biology and develop new therapeutic tools, especially immunoprophylactics.

We thank Drs S. Aaronson, F. Brun-Vézinet, R. Desrosiers, P. Fultz, H. Marquardt and M. Muesing for exchange of scientific information; Drs S. Wain-Hobson, M. L. Michel and A. Louise for providing plasmids, Bernard Caudron for help in preparation of the computer printouts; Dr. Lautaro Perez for discussion and encouragement. M.E. is a fellow of the Leukemia Society of America, Inc.

Received 4 March; accepted 19 March 1987

1. Weiss, R. A. in *RNA Tumor Viruses* (eds Weiss, R., Teich, N. Varmus, H. & Coffin, J.) Vol II, 405-485 (Cold Spring Harbor Laboratory, New York, 1985).
2. Curran, J. W. *et al. Science* 229, 1352-1357 (1985).
3. Quinn, T. C. Mann, J. M., Curran, J. W. & Piot, P. *Science* 234, 955-963 (1986).
4. Piot, P. *et al. Lancet* ii, 65-69 (1984).
5. Clumeck, N. *et al. N. Engl. J. Med.* 310, 492-497 (1984).
6. Barré-Sinoussi, F. *et al. Science* 220, 868-870 (1983).
7. Popovic, M., Samgadharan, M. G., Read, E. & Gallo, R. C. *Science* 224, 497-500 (1984).
8. Levy, J. A. *et al. Science* 225, 840-842 (1984).
9. Brun-Vezinet, F. *et al. Science* 226, 453-456 (1984).
10. Elfrodt, A. *et al. Lancet* i, 1383-1385 (1984).
11. Bailey, A. C. *et al. Lancet* i, 359-361 (1985).
12. Alizon, M., Wain-Hobson, S., Montagnier, L. & Sonigo, P. *Cell* 46, 63-74 (1986).
13. Wilkey, R. L. *et al. Proc. natn. Acad. Sci. U.S.A.* 83, 5038-5042 (1986).
14. Starcich, B. R. *et al. Cell* 45, 637-648 (1986).
15. Wain-Hobson, S. Sonigo, P., Danos, O., Cole, S. & Alizon, M. *Cell* 40, 9-17 (1985).
16. Sanchez Pescador, R. *et al. Science* 227, 484-492 (1985).
17. Ratner, L. *et al. Nature* 313, 277-284 (1985).
18. Muesing, M. A. *et al. Nature* 313, 450-458 (1985).
19. Clavel, F. *et al. Science* 233, 343-346 (1986).
20. Brun-Vézinet, F. *et al. Lancet* i, 128-132 (1987).
21. Clavel, F. *et al. N. Engl. J. Med.* (in the press).
22. Clavel, F. *et al. Nature* 324, 691-695 (1986).
23. Daniel, M. D. *et al. Science* 228, 1199-1204 (1985).
24. Kanki, P. J., Alroy, J. & Essex, M. *Science* 230, 951-954 (1985).
25. Benveniste, R. E. *et al. J. Virol.* 60, 483-490 (1986).
26. Murphey-Corb, M. *et al. Nature* 321, 435-437 (1986).
27. Sonigo, P. *et al. Cell* 42, 369-382 (1985).
28. Arya, S. K., Guo, C., Josephs, S. F. & Wong-Staal, F. *Science* 229, 69-73 (1985).
29. Sodroski, J., Patarca, R., Rosen, C., Wong-Staal, F. & Haseltine, W. *Science* 229, 74-77 (1985).
30. Sodroski, J. *et al. Nature* 321, 412-417 (1986).
31. Feinberg, M. B., Jarrett, R. F., Aldovini, A., Gallo, R. C. & Wong-Staal, F. *Cell* 46, 807-817 (1986).
32. Jones, K., Kadonaga, J., Luciw, P. & Tjian, R. *Science* 232, 755-759 (1986).
33. Muesing, M., Smith, D. H. & Capon, D. J. *Cell* 48, 691-701 (1987).
34. Robson, A. B. *et al. Science* 229, 1388-1390 (1985).
35. Lightfoot, M. *et al. J. Virol.* 60, 771-775 (1986).
36. Allan, J. S. *et al. Science* 228, 1091-1094 (1985).
37. Lee, T. H. *et al. Science* 231, 1546-1549 (1986).
38. Kan, N. C. *et al. Science* 231, 1553-1555 (1986).
39. Allan, J. S. *et al. Science* 230, 810-813 (1985).
40. Miller, J., Mac Laehlan, A. & Klug, A. *EMBO J.* 4, 1609-1614 (1985).
41. Dayton, A., Sodroski, J. G., Rosen, C. A., Goh, W. C. & Haseltine, W. A. *Cell* 44, 941-947 (1986).
42. Hess, J. L., Pyper, J. M. & Clements, J. E. *J. Virol.* 60, 385-393 (1986).
43. Gorman, C. M., Merlino, G. T., Willingham, M. C., Pastan, I. & Howard, B. H. *Proc. natn. Acad. Sci. U.S.A.* 79, 6777-6781 (1982).
44. Adachi, A. *et al. J. Virol.* 59, 284-291 (1986).
45. Rosen, C. A., Sodroski, J. G. & Haseltine, W. A. *Cell* 41, 813-823 (1985).
46. Coffin, J. M. *et al. Nature* 321, 10 (1986).
47. Kreiss, J. K. *et al. N. Engl. J. Med.* 314, 414-418 (1986).
48. Kanki, P. J. *et al. Science* 232, 238-243 (1986).
49. Bann, F. *et al. Lancet* ii, 1387-1389 (1985).
50. Dalgleish, A. G. *et al. Nature* 312, 763-767 (1984).
51. Klatzmann, D. *et al. Nature* 312, 767-768 (1984).
52. McDougal, J. S. *et al. Science* 231, 382-385 (1985).
53. Maddon, P. J. *et al. Cell* 47, 333-348 (1986).
54. Weiss, R. A. *et al. Nature* 316, 69-72 (1985).
55. Clavel, F., Klatzman, D. & Montagnier, L. *Lancet* i, 878-880 (1985).
56. Lasky, L. A. *et al. Science* 223, 209-212 (1985).
57. Messing, J. & Viera, J. *Gene* 19, 269-276 (1982).
58. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* 74, 5463-5467 (1977).
59. Gorman, C. M., Moffat, L. F. & Howard, B. H. *Molec. Cell. Biol.* 2, 1044-1051 (1982).
60. Graham, F. L. & Van der Eb, A. J. *Virology* 52, 456-467 (1973).
61. Wilbur, W. J. & Lipman, D. J. *Proc. natn. Acad. Sci. U.S.A.* 80, 726-730 (1983).
62. Seiki, M., Hatton, S., Hirayama, Y. & Yoshida, M. *Proc. natn. Acad. Sci. U.S.A.* 80, 3618-3622 (1983).
63. Kiyokawa, T., Yoshikura, H., Hatton, S., Seiki, M. & Yoshida, M. *Proc. natn. Acad. Sci. U.S.A.* 81, 6202-6206 (1984).
64. Sagata, N. *et al. Proc. natn. Acad. Sci. U.S.A.* 82, 677-681 (1985).
65. Schwartz, D. E., Tizard, R. & Gilbert, W. *Cell* 32, 853-869 (1983).

LETTERS TO NATURE

Switching phenomena in a new 90-K superconductor

P. H. Hor*, R. L. Meng*, C. W. Chu*, M. K. Wu†, E. Zirngiebl‡, J. D. Thompson‡ & C. Y. Huang§

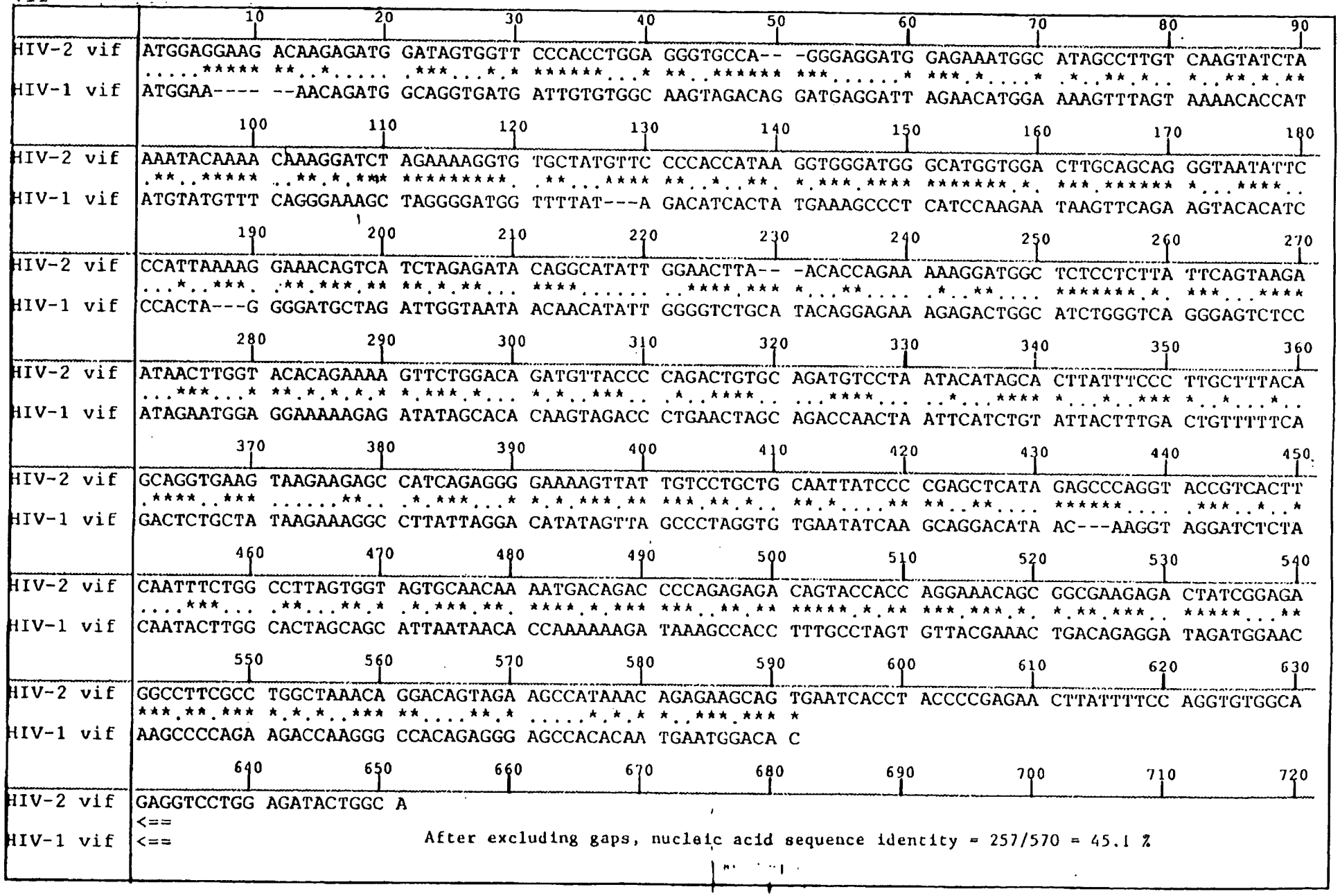
* Department of Physics, University of Houston, Houston, Texas 77004, USA
† Department of Physics, University of Alabama, Huntsville, Alabama 35899, USA
‡ Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA
§ Lockheed Palo Alto Research Laboratory, Palo Alto, California 94304-1191, USA

Recently, Wu *et al.*[1] and Hor *et al.*[2] have shown that $Y_{1.2}Ba_{0.8}CuO_{4-\delta}$ is a superconductor with a superconducting onset temperature at ~92 K as determined by their resistivity and a.c. susceptibility measurements. Because the magnetic properties are important in describing the nature of superconductivity, we have measured the d.c. magnetic moment of this material. Here we show that this material cooled in zero field or in a high field (H_{cool} > 90 G) is diamagnetic below $T_{cm} \approx 90$ K, consistent with the previous measurements[1,2]. However, when the sample is cooled in a small field (≤85 G), the magnetization, M, first becomes negative (diamagnetic) below T_{cm}, but further cooling results in a jump of M to a positive value at low temperature. We have also observed this switching by the application of an additional small field when the sample was cooled in a small field.

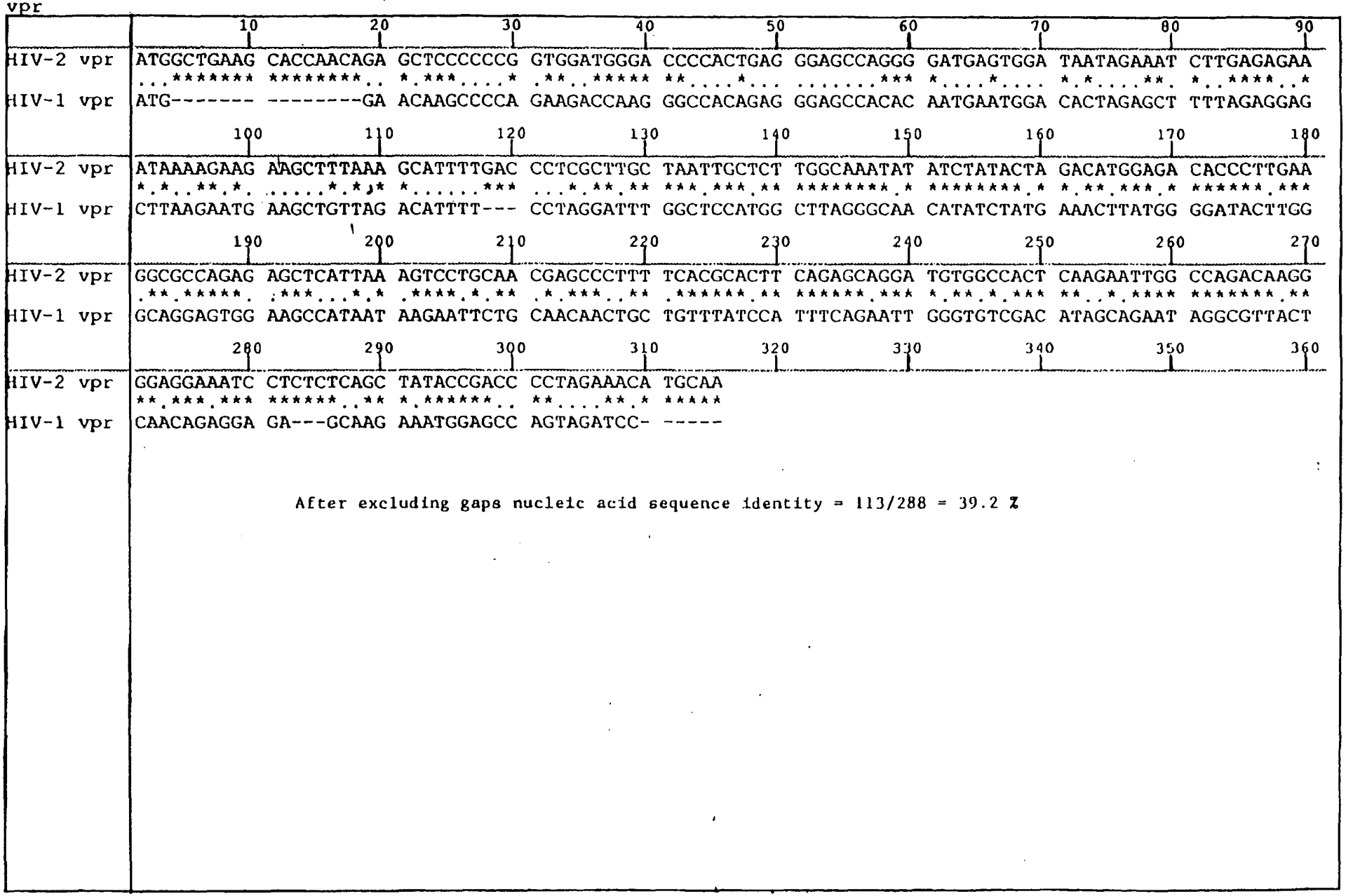
The $Y_{1.2}Ba_{0.8}CuO_{4-\delta}$ sample was prepared as described in ref. 1. The X-ray diffractograms reveal that the sample has multiple phases, devoid of the K_2NiF_4 structure. From the electrical resistance measurement, the superconducting onset temperature is $T_{co} \approx 94.5$ K and the resistance becomes 'zero' below $T_0 \approx 92$ K indicating that the sample is a superconductor with a rather narrow transition width. A Quantum Design superconducting quantum interference device (SQUID) magnetometer has been employed to measure the magnetization of the sample as a function of temperature and magnetic field. When the sample is cooled under zero field conditions, we have found that M is diamagnetic below T_{cm} and the susceptibility below ~25 K reaches ~35% of that of perfect diamagnetism ($-1/4\pi$).

We have also measured M when the sample is cooled in a field, H_{cool}. In Fig. 1, the magnetization obtained at various

vif

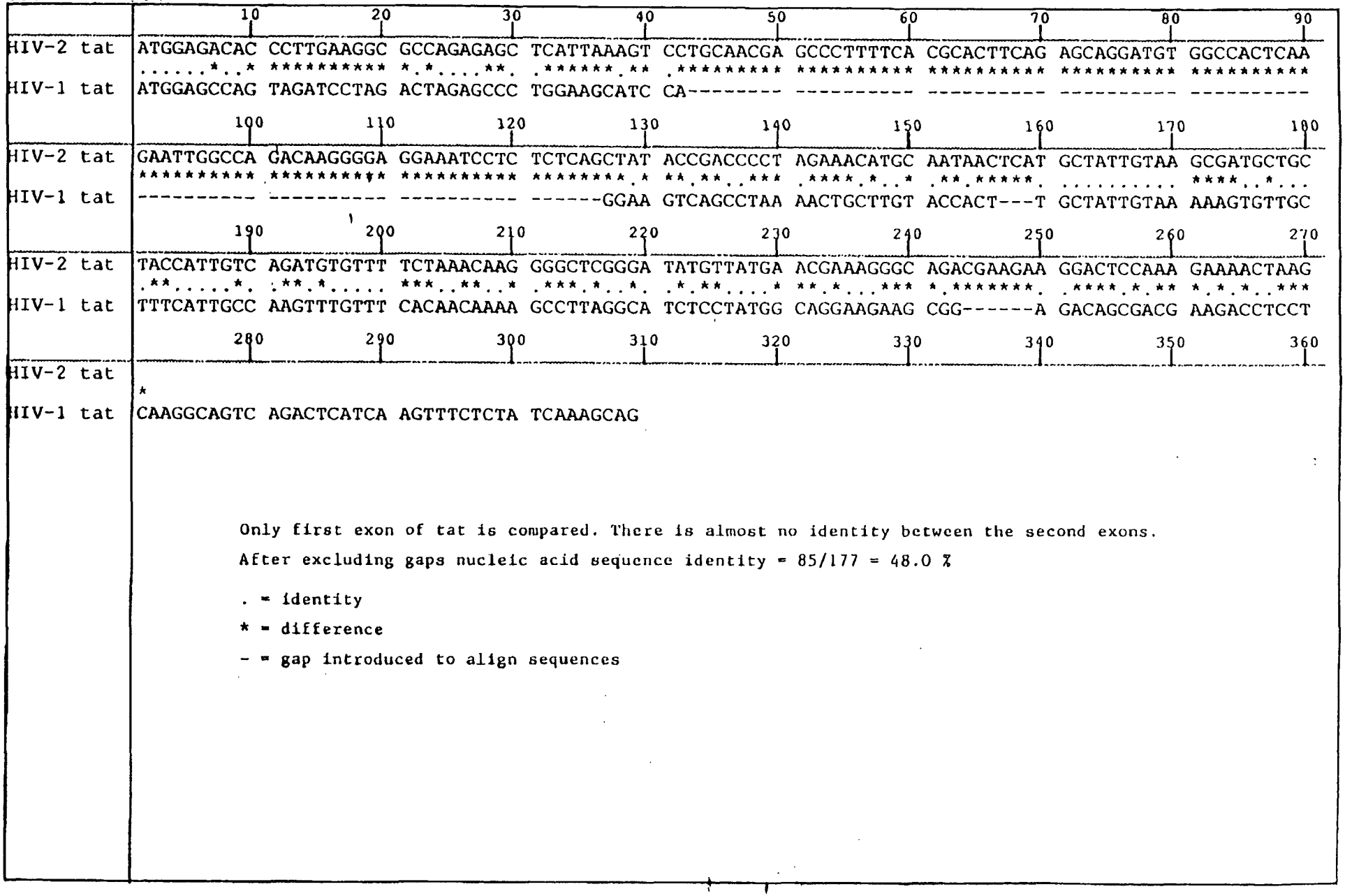

10 20 30 40 50 60 70 80 90
HIV-2 vif ATGGAGGAAG ACAAGAGATG GATAGTGGTT CCCACCTGGA GGGTGCCA-- -GGGAGGATG GAGAAATGGC ATAGCCTTGT CAAGTATCTA
HIV-1 vif ATGGAA---- --AACAGATG GCAGGTGATG ATTGTGTGGC AAGTAGACAG GATGAGGATT AGAACATGGA AAAGTTTAGT AAAACACCAT
100 110 120 130 140 150 160 170 180
HIV-2 vif AAATACAAAA CAAAGGATCT AGAAAAGGTG TGCTATGTTC CCCACCATAA GGTGGGATGG GCATGGTGGA CTTGCAGCAG GGTAATATTC
HIV-1 vif ATGTATGTTT CAGGGAAAGC TAGGGGATGG TTTTAT---A GACATCACTA TGAAAGCCCT CATCCAAGAA TAAGTTCAGA AGTACACATC
190 200 210 220 230 240 250 260 270
HIV-2 vif CCATTAAAAG GAAACAGTCA TCTAGAGATA CAGGCATATT GGAACTTA-- -ACACCAGAA AAAGGATGGC TCTCCTCTTA TTCAGTAAGA
HIV-1 vif CCACTA---G GGGATGCTAG ATTGGTAATA ACAACATATT GGGGTCTGCA TACAGGAGAA AGAGACTGGC ATCTGGGTCA GGGAGTCTCC
280 290 300 310 320 330 340 350 360
HIV-2 vif ATAACTTGGT ACACAGAAAA GTTCTGGACA GATGTTACCC CAGACTGTGC AGATGTCCTA ATACATAGCA CTTATTTCCC TTGCTTTACA
HIV-1 vif ATAGAATGGA GGAAAAAGAG ATATAGCACA CAAGTAGACC CTGAACTAGC AGACCAACTA ATTCATCTGT ATTACTTTGA CTGTTTTTCA
370 380 390 400 410 420 430 440 450
HIV-2 vif GCAGGTGAAG TAAGAAGAGC CATCAGAGGG GAAAAGTTAT TGTCCTGCTG CAATTATCCC CGAGCTCATA GAGCCCAGGT ACCGTCACTT
HIV-1 vif GACTCTGCTA TAAGAAAGGC CTTATTAGGA CATATAGTTA GCCCTAGGTG TGAATATCAA GCAGGACATA AC---AAGGT AGGATCTCTA
460 470 480 490 500 510 520 530 540
HIV-2 vif CAATTTCTGG CCTTAGTGGT AGTGCAACAA AATGACAGAC CCCAGAGAGA CAGTACCACC AGGAAACAGC GGCGAAGAGA CTATCGGAGA
HIV-1 vif CAATACTTGG CACTAGCAGC ATTAATAACA CCAAAAAAGA TAAAGCCACC TTTGCCTAGT GTTACGAAAC TGACAGAGGA TAGATGGAAC
550 560 570 580 590 600 610 620 630
HIV-2 vif GGCCTTCGCC TGGCTAAACA GGACAGTAGA AGCCATAAAC AGAGAAGCAG TGAATCACCT ACCCCGAGAA CTTATTTTCC AGGTGTGGCA
HIV-1 vif AAGCCCCAGA AGACCAAGGG CCACAGAGGG AGCCACACAA TGAATGGACA C
640 650 660 670 680 690 700 710 720
HIV-2 vif GAGGTCCTGG AGATACTGGC A
HIV-1 vif
After excluding gaps, nucleic acid sequence identity = 257/570 = 45.1 %

vpr



After excluding gaps nucleic acid sequence identity = 113/288 = 39.2 %

tat



Only first exon of tat is compared. There is almost no identity between the second exons.

After excluding gaps nucleic acid sequence identity = 85/177 = 48.0 %

. = identity

* = difference

- = gap introduced to align sequences

Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 2209–2213, April 1986
Medical Sciences

Three novel genes of human T-lymphotropic virus type III: Immune reactivity of their products with sera from acquired immune deficiency syndrome patients

(*sor*, *tat* and *3' orf* genes/cDNA cloning/double splicing/*in vitro* translation/immunoprecipitation)

SURESH K. ARYA* AND ROBERT C. GALLO

Laboratory of Tumor Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, MD 20205

Communicated by Peter C. Nowell, November 13, 1985

ABSTRACT **Human T-lymphotropic virus type III or lymphoadenopathy associated virus (HTLV-III/LAV) is the cause of acquired immune deficiency syndrome (AIDS). In addition to the conventional retroviral genes involved in virus replication, namely, *gag*, *pol*, and *env* genes, DNA sequence analysis of HTLV-III genome predicted two additional open reading frames termed by us short open reading frame (*sor*) and 3' open reading frame (*3' orf*). Furthermore, functional analysis revealed another gene with transactivating function, termed *tat*. We have now structurally identified and functionally characterized these HTLV-III specific genes by way of cDNA cloning. DNA sequence analysis of the clones shows that the *tat* and *3' orf* genes contain three exons and their transcription into functional mRNA involves two splicing events and that the *sor* gene contains at least two exons. *In vitro* transcription and translation of the cloned spliced sequences show that the *sor*, *tat*, and *3' orf* genes code for polypeptides with apparent mobility of 24–25 kDa, 14–15 kDa, and 26–28 kDa, respectively. All three polypeptides are immune reactive and are immunogenic in the natural host. The results demonstrate that the three extra open reading frames of HTLV-III, two of which are unique to HTLV-III, are in fact genes that function *in vivo* and further allow the identification of three new and previously unrecognized HTLV-III antigens with differential immunogenicity in individuals with acquired immune deficiency syndrome and related disorders.**

Human T-lymphotropic virus type III (HTLV-III) or the lymphoadenopathy associated virus (LAV) is etiologically linked to acquired immune deficiency syndrome (AIDS) and AIDS-related complex (ARC) (1–4). The overall genetic structure of HTLV-III/LAV is similar to that of other animal retroviruses. However, besides *gag*, *pol*, and *env* genes, DNA sequence analysis of HTLV-III/LAV genome predicted two additional open reading frames or potential genes (5–8), termed by us and others *sor* (short open reading frame) and *3' orf* (3' open reading frame). The presence of a third gene, termed *tat* (transactivation of transcription), was also suggested (9–12). Thus, HTLV-III/LAV contains coding potential for three genes that are specific to this virus. Two of these putative genes, *sor* and *3' orf*, are unique to HTLV-III but a functional analog of the third gene, *tat*, is also carried by other members of the HTLV-bovine leukemia virus (BLV) group of retroviruses (9–13). We and others have localized the *tat* gene of HTLV-III to a region between the putative *sor* and the *env* genes (10, 12), a region of the genome previously thought to be noncoding. This is distinct from the other members of the HTLV-BLV group where *tat* gene is located downstream from the *env* gene. Thus, even the *tat* gene is organized differently in HTLV-III. We report here that *sor*, *tat*, and *3' orf* genes all contain intron(s) and are respectively translated into polypeptides with apparent mobility of 24–25 kDa, 14–15 kDa, and 26–28 kDa on $NaDodSO_4$/PAGE. These gene products display differential immune reactivity for HTLV-III positive human sera, the *3' orf* gene product being the most immune reactive. The results demonstrate the existence of three new HTLV-III antigens.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

cDNA Cloning and DNA Sequencing. Poly(A)-selected RNA from HTLV-III-infected H4 cells, isolated as described (14, 15), was used to construct cDNA libraries as reported (10). The libraries were screened with subgenomic HTLV-III probes to obtain clones containing specific HTLV-III sequences (10, 11). The selected clones were characterized by restriction mapping and DNA sequencing by the method of Maxam and Gilbert (16).

***In Vitro* Transcription and Translation.** The inserts of selected cDNA clones were transferred to the vector pSP6 that transcribes inserted DNA under the influence of SP6 promoter (17). RNA was transcribed *in vitro* after linearization of the plasmid DNA with specific restriction enzymes. It was translated *in vitro* by using rabbit reticulocyte translation system and [^{35}S]methionine, and the products were analyzed by 12% $NaDodSO_4$/PAGE and radioautography by the standard procedures.

Immunoprecipitation with Human Sera. The *in vitro* translation products were incubated with normal human serum for 1–2 hr at 4°C. Suspension of *Staphylococcus aureus* (Staph A) cells was then added, and incubation was continued for an additional 1 hr. The sample was centrifuged, and the supernatant was divided into two equal parts, one of which was incubated with immune serum at 4°C for 18–24 hr. A suspension of Staph A was added to each sample and incubated at 4°C for 1 hr. The samples were centrifuged, and the pellets were repeatedly and sequentially washed with 50 mM Tris·HCl (pH 7.4)/50 mM EDTA/0.05% Nonidet P-40/1% aproteinin containing 0.5 M NaCl or 0.15 mM NaCl. The pellets were suspended in 75 mM Tris·HCl (pH 6.8)/0.7 mM 2-mercaptoethanol/2% (wt/vol) $NaDodSO_4$/10% (vol/vol) glycerol/0.001% bromophenol blue, boiled for 10 min, and centrifuged. The supernatants were subjected to 12% $NaDodSO_4$/PAGE analysis.

RESULTS

cDNA Clones of HTLV-III Specific Genes. To identify HTLV-III specific genes, we took the direct approach of

Abbreviations: HTLV-III/LAV, human T-lymphotropic virus type III/lymphoadenopathy associated virus; AIDS, acquired immune deficiency syndrome; bp, base pair(s); kb, kilobase(s).

*To whom correspondence should be addressed.

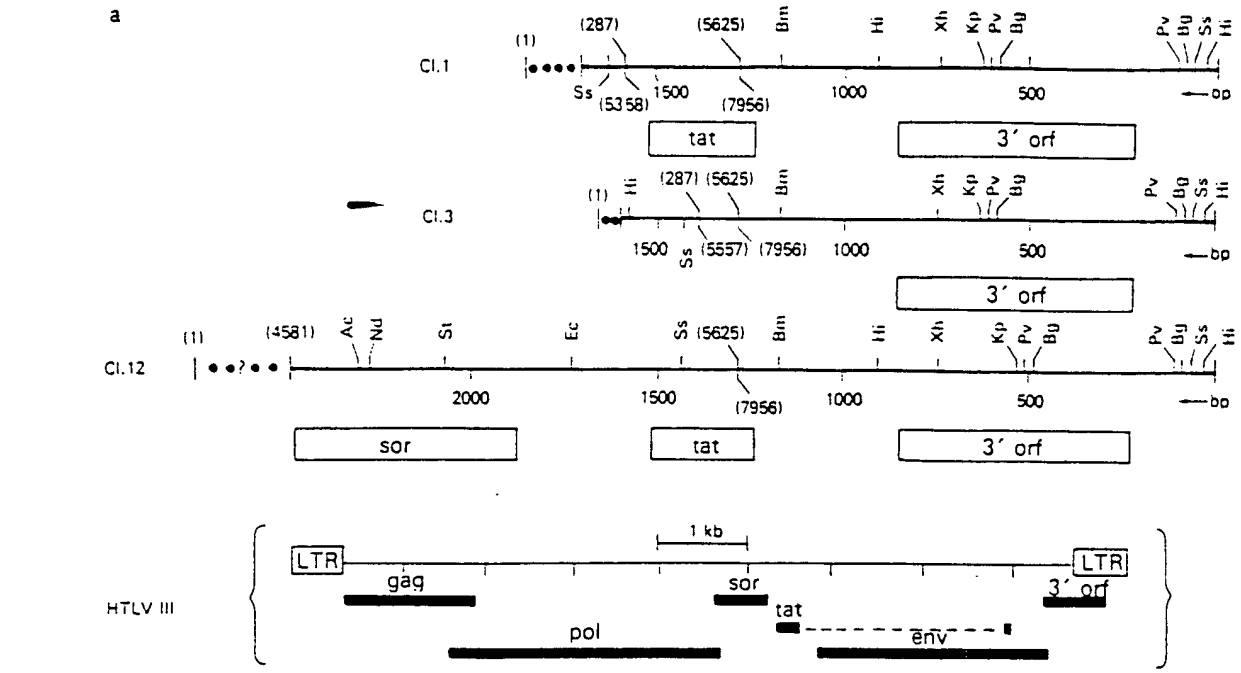


FIG. 1. *(Legend appears at the bottom of the opposite page.)*

obtaining functional cDNA clones by screening cDNA libraries with specific subgenomic HTLV-III probes to obtain the desired clones. We have previously described a functional cDNA clone (clone 1) corresponding to the mRNA of the *tat* gene (10). We also described in this previous report a second cDNA clone (clone 3) that we speculated may correspond to the mRNA of *3' orf* gene. Both of these clones were copies of the mRNAs that were generated by double splicing events. Thus, the *tat* gene and the putative *3' orf* gene consisted of three exons and two introns (Fig. 1). We have now obtained another cDNA clone (clone 12) that contains the complete open reading frame of the putative *sor* gene, in addition to the open reading frames of the *tat* gene and the putative *3' orf* gene (Fig. 1). DNA sequence analysis of clone 12 [2304 base pairs (bp)] showed it to be an incomplete cDNA clone as it lacked the mRNA cap site and possibly other sequences on the 5' side of the *sor* open reading frame. However, it contained the 3'-splice junction that was identical to the 3'-splice junction of clones 1 and 3 (Fig. 1).

Translation Products of HTLV-III Specific Genes. To characterize the gene products of the putative *sor*, *tat*, and *3' orf* genes, the cDNA were transferred to the transcription vector pSP6, and the plasmids containing clones 1, 3, and 12 cDNA inserts were designated pSP-1, pSP-3, and pSP-12, respectively (Fig. 2*a*). RNA was transcribed after linearization of the plasmid DNAs with specific restriction enzymes that were chosen because they will either retain a given open reading frame as a part of pSP6 transcriptional unit or delete it. The transcription of the plasmid DNAs cleaved with specific restriction enzymes gave RNA transcripts of the appropriate sizes (data not shown). These transcripts were translated and products analyzed. Representative results are shown in Fig. 2. The transcripts of pSP-1 DNA linearized with *Xba* I or *Sma I* gave two polypeptides with apparent mobility of 25–26 kDa and 14–15 kDa, the 14–15 kDa polypeptide being in smaller relative amounts. Digestion of this plasmid DNA with *Bam*HI or *Xho* I, which deletes *3' orf* open reading frame from the transcriptional unit, gave only the polypeptide with 14–15 kDa apparent mobility. These results suggest that 25–26 kDa and 14–15 kDa polypeptides were products of the *3' orf* and *tat* open reading frames, respectively. While pSP-12 DNA linearized with *Xba* I displayed three polypeptides of 25–26 kDa, 23–24 kDa, and 14–15 kDa, this DNA linearized with *Bam*HI or *Xho* I gave only two polypeptides of 23–24 kDa and 14–15 kDa apparent mobility (Fig. 2). These results again suggest that 25–26 kDa and 14–15 kDa polypeptides are the product of the *3' orf* and *tat* open reading frames, respectively, and further suggest that 23–24 kDa polypeptide is the product of the *sor* open reading frame.

The transcripts of *Xba* I linearized pSP-3 DNA, which contains only the *3' orf* open reading frames, though not always translated efficiently, displayed a distinct polypeptide with apparent mobility of 27–28 kDa. This polypeptide was not detected when pSP-1 DNA was linearized with *Bam*HI or *Xho* I, which removes the *3' orf* open reading frame from the transcriptional unit. These results suggest that the *3' orf* open reading frame contained in pSP-3 DNA was being translated into a 27–28 kDa polypeptide. The plasmid DNAs containing cDNA inserts in the incorrect orientation with respect to the SP6 promoter gave transcripts of the appropriate sizes but none of these transcripts were translated into distinct polypeptides (Fig. 2).

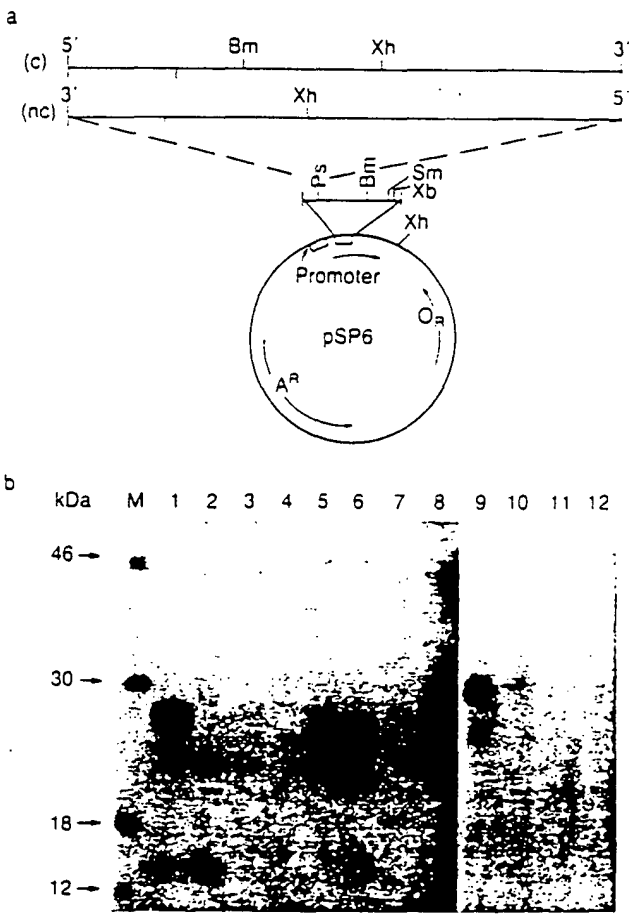


FIG. 2. (*a*) Physical map of pSP-1 containing HTLV-III cDNA clone 1. pSP-3 and pSP-12 were similarly constructed. (c) and (nc) refer to the correct and noncorrect orientation of the cDNA insert with respect to the SP6 promoter. (*b*) $NaDodSO_4$/PAGE analysis of the translation products of the transcripts from pSP-1, pSP-12, and pSP-3 plasmid DNAs. Lanes 1 and 2, *Xba* I- and *Bam*HI-digested pSP-1(c) DNA; lanes 3 and 4, *Xba* I- and *Bam*HI-digested pSP-1(nc) DNA; lanes 5 and 6, *Xba* I- and *Bam*HI-digested pSP-12(c) DNA; lanes 7 and 8, *Xba* I- and *Bam*HI-digested pSP-12(nc) DNA; lanes 9 and 10, *Xba* I- and *Bam*HI-digested pSP-3(c) DNA; lanes 11 and 12, *Xba* I- and *Bam*HI-digested pSP-3(nc) DNA; lane M, molecular size standards.

Since clone 12 contained the *tat* open reading frame in addition to the *sor* and *3' orf* open reading frames, we tested its transactivating capacity in a transfection system that measures transactivation of the chloramphenicol acetyl transferase (CAT) gene (see ref. 10). Representative results for human lymphoid JM cells are shown in Fig. 3. Clearly, clone 12 DNA transactivated the *CAT* gene activity. Thus, the *tat* open reading frame contained in clone 12 was transcribed and translated into a functionally active polypeptide.

Immune Reactivity of HTLV-III Specific Gene Products. To evaluate the immune reactivity of the polypeptides directed by the *sor*, *tat*, and *3' orf* open reading frames, translation products were immune precipitated with HTLV-III-positive human sera from several individuals. Representative results are shown in Figs. 4 and 5, and data are compiled in Table 1. HTLV-III-positive serum specifically immune precipitated a

FIG. 1 (*on opposite page*). (*a*) Physical maps of HTLV-III cDNA clones 1, 3, and 12. The two splice junctions for clones 1 and 3 and one splice junction for clone 12 are indicated. The nucleotide numbering in parentheses is according to Ratner *et al.* (6). (*b*) DNA sequence of HTLV-III cDNA clone 12. The three open reading frames contained in this clone along with the predicted amino acid sequences are shown. DNA sequences of clones 1 and 3 have been reported before (10). The open reading frame for the *tat* gene is in a different frame than those of the *sor* and *3' orf* genes.

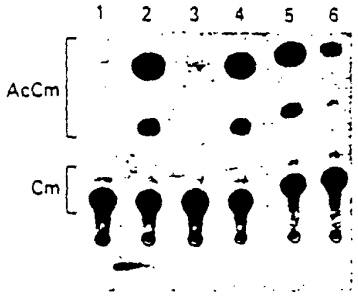


FIG. 3. Enhancement of HTLV-III LTR-promoted *CAT* gene expression by HTLV-III cDNA clone 12. Human lymphoid JM cells were cotransfected with clone 12 DNA in expression vector pCV (pCV-12) and HTLV-III LTR-CAT (pC15-CAT) plasmid DNA by the DEAE-dextran protocol (10). *CAT* gene product activity in the extract of transfected cells was measured by analyzing the conversion of [^{14}C]chloramphenicol (Cm) into its acetylated forms (AcCm) by thin layer paper chromatography. Lanes 1 to 6 are respectively for cells transfected with DNAs of pSVoCAT (1), pRSVCAT (2), pC15-CAT (3), pC15-CAT plus pEV-HXb3 (4), pC15-CAT plus pCV-12 (correct orientation) (5), and pC15-CAT plus pCV-12 (incorrect orientation) (6).

predominant polypeptide of 25–26 kDa for *Xba* I as well as *Sma* I linearized plasmid pSP-1 DNA. The 25–26 kDa polypeptide was also specifically immune precipitated from the translation products of *Xba* I as well as *Sma* I linearized plasmid pSP-12 DNA. Similar results were obtained with plasmid pSP-3 DNA, except the apparent size of this polypeptide was 27–28 kDa (Fig. 4). The marginal detection of this polypeptide in translation products of *Bam*HI-digested pSP-1 and pSP-3 plasmid DNAs was probably the result of incomplete enzyme digestions; it was not detected for *Bam*HI-digested pSP-12 plasmid DNA. Instead, translation products of *Bam*HI-digested pSP-12 plasmid DNA displayed a band at 23–24 kDa that was immune precipitated with HTLV-III-positive serum but also to a lesser extent with some normal human sera (see Table 1). Consistent with our interpretation of the translation products noted above, we infer that the 26–28 kDa and 23–24 kDa polypeptides are the immune reactive products of *3' orf* and *sor* open reading frames, respectively. Immunoprecipitation of the 14–15 kDa *tat* gene product was not obvious with this particular HTLV-III-positive serum but could be detected to varying extent by some of the other HTLV-III-positive sera as shown in Fig. 5 and listed in Table 1.

DISCUSSION

The HTLV-III open reading frames termed *sor*, *tat*, and *3' orf* are specific to this virus and two of these, *sor* and *3' orf*, are unique to it. The products of the three HTLV-III specific open reading frames immune react with antibodies in sera of individuals with AIDS and ARC. Therefore, these open reading frames are in fact genes that are expressed *in vivo*.

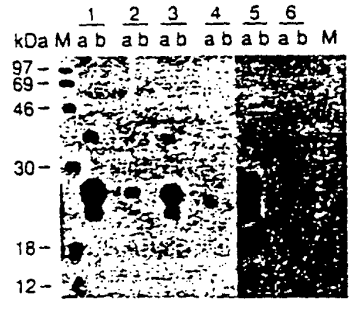


FIG. 4. NaDodSO$_4$/PAGE analysis of immune precipitates of translation products of pSP-1, pSP-12, and pSP-3 DNA transcripts. Lanes 1 and 2, *Xba* I- and *Bam*HI-digested pSP-1 DNA; lanes 3 and 4, *Xba* I- and *Bam*HI-digested pSP-12 DNA; lanes 5 and 6, *Xba* I- and *Bam*HI-digested pSP-3 DNA. Sublanes (a) and (b) are for HTLV-III-positive and normal human serum, respectively. Lane M, molecular size standards.

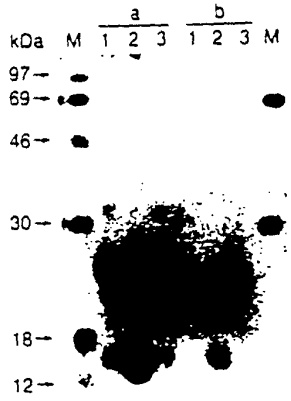


FIG. 5. NaDodSO$_4$/PAGE analysis of immune precipitates of translation products of *Xba* I-digested pSP-1 (*a*) and pSP-12 DNA (*b*) transcripts. Lanes 1 and 2 are for two different immune sera and lane 3 is for normal serum. Lane M, molecular size standards. Analysis was performed as described in Fig. 3. (The 25–26-kDa band is the *3' orf* gene product, presumably the result of incomplete enzyme digestion of the plasmid DNA.)

Our results allow the structural definition of the three HTLV-III specific genes. We have previously characterized the functional domain of the *tat* gene (10). Like the *tat* gene, the *3' orf* gene (clone 3) also consists of three exons (287 bp, 69 bp, and 1258 bp) and two introns (5268 bp and 2330 bp), and its transcription into a functional mRNA involves double splicing. The *3' orf* gene differs from the *tat* gene in having a truncated second exon involving splicing out of the putative initiation codon of the *tat* gene product (10).

The *sor* gene contains at least two and probably three exons. The 3' exon (1258 bp) of this gene (clone 12) is identical to the third exon of the *tat* and *3' orf* genes. The sequences on the 5' side (1114 bp) of this exon in clone 12 are shared with the second exon of the *tat* gene and extend upstream to include *sor* open reading frame. We suspect that the generation of the *sor* mRNA also involves two splicing events. It is possible that the synthesis of this mRNA involves the same first donor site (at nucleotide 287) as other mRNAs and one of the many potential acceptor sites located to the 5' side of the *sor* open reading frame. If the consensus acceptor site nearest to the 5' side of the *sor* open reading frame located at nucleotide 4494 is utilized, the functional *sor* gene will generate a message of about 2.7 kilobases (kb). However, if the *sor* message involves only one splicing event demonstrated in clone 12, the mRNA would be about 7.0 kb.

Table 1. Immune reactivity of the *sor*, *tat*, and *3' orf* gene products

Serum		Gene product		
Number	Diagnosis	*sor*	*tat*	*3' orf*
1	AIDS	+	±	++
2	AIDS	+	+	++
3	AIDS	+	±	++
4	AIDS	±	—	±
5	ARC	+	++	++
6	ARC	+	—	++
7	Healthy homosexual	—	—	—
8	Healthy homosexual	—	—	++
9	Healthy homosexual	—	—	—
10	Healthy homosexual	—	—	—
11	Healthy heterosexual	±	—	—
12	Healthy heterosexual	+	—	—
13	Healthy heterosexual	—	—	—
14	Healthy heterosexual	—	—	—

+, Reactive; ±, detectable; ++, strongly reactive; —, not detected.

Muesing *et al.* (8) have suggested that the *sor* gene consists of two exons generating a message of about 5.0 kb. Their suggestion is inconsistent with clone 12 that contains an intron located within their suggested second exon. It is possible to postulate other combinations of potential 5' donor and acceptor splice sites to generate a 5-kb message involving double splicing. It is, of course, possible that more than one species of the *sor* mRNA is synthesized utilizing alternative splicing events. We have previously reported four abundant mRNAs of 9.4 kb, 4.2 kb, 2.0 kb, and 1.8 kb in HTLV-III-infected cells (10, 11). We also observed other less abundant RNA species of about 7 kb, 5 kb, 3.2 kb, and 2.8 kb in these cells. One or more of these species could correspond to the *sor* message.

The *sor*, *tat*, and *3' orf* genes synthesize polypeptides with apparent mobilities of 23–24 kDa, 14–15 kDa, and 26–28 kDa, respectively. The *3' orf* open reading in clone 3 and in clones 1 and 12 was translated into a polypeptide of 27–28 kDa and 25–26 kDa, respectively. This open reading frame contains two initiation codons (ATG) 57 bp apart in phase in its 5' portion (Fig. 1). We suggest that the first and second ATGs are used for translation in pSP-3 DNA and pSP-1 and pSP-12 DNAs, respectively. Both of these ATG triplets are flanked by the appropriate consensus sequence requisite for efficient translation initiation by the eukaryotic ribosomes (18). Furthermore, the coding potential of the open reading frames for the *sor*, *tat*, and *3' orf* genes, starting from the first in phase initiation codon is respectively 192, 86, and 206 amino acid residues, predicting the respective polypeptides of about 20 kDa, 9 kDa, and 21 kDa. The observed mobility of the products of these genes in $NaDodSO_4$/PAGE was uniformly higher than predicted. This may suggest anomalous conformation and/or posttranslational modifications of the proteins.

The products of the *sor*, *tat*, and *3' orf* genes are immunogenic *in vivo*, thus identifying three new antigens for HTLV-III, in addition to the previously described *gag* and *env* gene products (19–22). The three gene products appear to be differentially immunoreactive and immunogenic, the *3' orf* gene product apparently being the most potent and the *sor* gene product being the least potent in this regard. The lesser immunogenicity of the *sor* gene product may be due to its diminished expression *in vivo* and its particular intracellular localization, or it may be related to its structure (Fig. 6). The predicted amino acid sequence of the *sor* gene product does not contain a cluster of amino acid residues that will impart to this protein hydrophilic structure with β-turns-two parameters generally thought to be responsible for strong immunogenicity (23, 24). Notably, the predicted amino acid sequence of both the *sor* and *tat* gene products lacks typical sequence (-aspargine-Xaa-threonine or serine-) that generally serves as a glycosylation site and such a sequence is present twice in the predicted sequence of *3' orf* gene product.

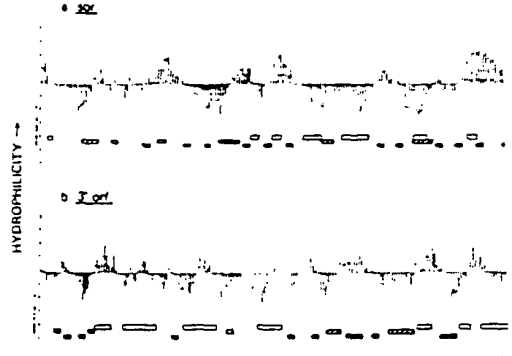


FIG. 6. Hydrophobicity profile and predicted secondary structure of *sor* (*a*) and *3' orf* (*b*) polypeptides analyzed according to Kyte and Doolittle (23) and Chou and Fasman (24). Secondary structure is depicted by boxes and vertical lines represent amino acid residues. Open box, α-helix; hatched box, β-sheets; closed box, β-turns.

With regard to any correlation between the progression of the disease and expression of the HTLV-III specific genes, the survey reported here is too small to detect meaningful trends. We think it is premature to draw conclusions from the observation that antibodies to the *3' orf* gene product were detected in all but one of the six sera from patients with AIDS and ARC but in only one out of four sera from HTLV-III-positive healthy homosexual individuals in this study. Further, some of the normal human sera reacted, though poorly, with the *sor* gene product. Although we cannot presently rule out artifactual interactions, this may suggest that a normal cellular gene with some homology to the *sor* gene exists, and its product is synthesized in some instances. The differential expression of the *sor*, *tat*, and *3' orf* genes *in vivo* may reflect mutual modulatory role(s) of the products of these genes.

We thank M. Sarngadharan, M. Guroff, and their colleagues and collaborators for providing serum samples used in this study. Additionally, thanks are due to L. Jagodzinski and R. Liou for assistance with DNA sequencing, and M. B. Eiden, C. Guo, and S. F. Josephs for useful discussions.

1. Barre-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rosenbaum, W. & Montagnier, L. (1983) *Science* **220**, 868–870.
2. Popovic, M., Sarngadharan, M. G., Read, E. & Gallo, R. C. (1984) *Science* **224**, 497–500.
3. Gallo, R. C., Salahuddin, S. Z., Popovic, M., Shearer, G. M., Kaplan, M., Haynes, B. F., Palker, T. J., Redfield, R., Oleske, J., Safai, B., White, C., Foster, P. & Markham, P. D. (1984) *Science* **224**, 500–503.
4. Hahn, B. H., Shaw, G. M., Arya, S. K., Popovic, M., Gallo, R. C. & Wong-Staal, F. (1984) *Nature (London)* **313**, 166–168.
5. Sanchez, R., Power, M. D., Barr, P. J., Steimer, K. S., Stempien, M. M., Brown-Shimer, S. L., Gee, W. W., Renard, A., Randolph, A., Levy, J. A., Dina, D. & Luciw, P. A. (1985) *Science* **227**, 484–492.
6. Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, B., Josephs, S. F., Doran, E. R., Rafalski, J. A., Whitehorn, E. A., Baumeister, K., Ivanoff, L., Petteway, S. R., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghrayeb, J., Chang, N. T., Gallo, R. C. & Wong-Staal, F. (1985) *Nature (London)* **313**, 277–284.
7. Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. & Alizon, M. (1985) *Cell* **40**, 9–17.
8. Muesing, M. A., Smith, D. H., Cabradilla, C. D., Benton, C. V., Lasky, L. A. & Capon, D. J. (1985) *Nature (London)* **313**, 450–458.
9. Sodroski, J., Rosen, C., Wong-Staal, F., Salahuddin, S. Z., Popovic, M., Arya, S. K., Gallo, R. C. & Haseltine, W. A. (1985) *Science* **227**, 171–173.
10. Arya, S. K., Guo, C., Josephs, S. F. & Wong-Staal, F. (1985) *Science* **229**, 69–73.
11. Arya, S. K., Gallo, R. C., Hahn, B. H., Shaw, G. M., Popovic, M., Salahuddin, S. Z. & Wong-Staal, F. (1984) *Science* **225**, 927–930.
12. Sodroski, J., Patarca, R., Rosen, C., Wong-Staal, F. & Haseltine, W. (1985) *Science* **229**, 74–77.
13. Sodroski, J. G., Rosen, C. A. & Haseltine, W. A. (1984) *Science* **225**, 381–385.
14. Arya, S. K. (1982) *Int. J. Biochem.* **14**, 19–24.
15. Arya, S. K. & Gallo, R. C. (1984) *Biochemistry* **23**, 6690–6696.
16. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
17. Melton, D., Krieg, P., Rebagliati, M., Maniatis, T., Zinn, K. & Green, M. (1984) *Nucleic Acids Res.* **12**, 7035–7056.
18. Kozak, M. (1984) *Nature (London)* **308**, 241–245.
19. Sarngadharan, M. G., Bruch, L., Popovic, M. & Gallo, R. C. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3481–3484.
20. Allan, J. S., Colligan, J. E., Barin, F., McLane, M. F., Sodroski, J. G., Rosen, C. A., Haseltine, W. A., Lee, T. H. & Essex, M. (1985) *Science* **228**, 1091–1094.
21. Barin, F., McLane, M. F., Allan, J. S. & Lee, T. H. (1985) *Science* **228**, 1094–1096.
22. Crowl, R., Ganguly, K., Gordon, M., Conroy, R., Schaber, M., Kramer, R., Shaw, G., Wong-Staal, F. & Reddy, E. P. (1985) *Cell* **41**, 979–986.
23. Kyte, J. & Doolittle, R. F. (1981) *J. Mol. Biol.* **157**, 105–132.
24. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13**, 222–228.

nature

11 August 1988
Vol. 334 Issue no. 6182

Inactivation of two genes in *Yersinia pseudotuberculosis* causes a significant increase in virulence and may explain, in part, the variations in virulence of *Yersinia pestis* that accounts for the rise and fall of plague ('black death') epidemics. See page 522 and News and Views. Cover shows 'St Charles Borromeo gives communion to plague victims', by Sigismondo Caula (E T Archive).

HIS WEEK

o cool to resist

bulk synthesis of a Tl–Ba–Cu–O superconductor with CuO_2 layers per unit cell is cribed on page 510. With a sition temperature of > 120 it continues the trend of easing T_c with the number CuO_2 layers. The race for erconductivity, Book Rev- , page 479.

agnetic bug

ewly isolated marine mag- otactic bacterium that un- pectedly synthesizes magnet-

in anaerobic conditions can ntribute to natural remanent agnetism found in long-term naerobic sediments, page 518.

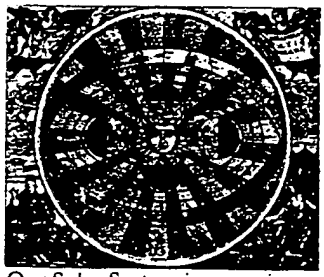
Dim sun

he eclipsing millisecond pulsar 57+20 has three candidates for companion. Optical studies ow the probable candidate is a riable object (star X) of low minosity, consistent with odels in which the pulsar wind is ainly in the form of low-energy rays or X-rays. Page 504.

alcium control

hree ionic mechanisms main- in elevated levels of intracel- lar calcium in mast cells and ay thus enhance calcium- ependent functions such as cretion, page 499.

Planets suite



Our Solar System is not unique: other stars in our Galaxy seem to have giant planets and new planetary systems are forming elesewhere. See pages 467 and 474.

Target practice

CD4-bearing T cells *in vitro* can capture, process and present gp120, rendering uninfected T cells a target for the anti-HIV T-cell response, page 530.

Take a neutrino . . .

Results from neutrino detectors confirm that too few neutrinos reach us from the Sun, calling for either new physics or new astronomy to provide an explanation. See Review Article.

Sink not source

Biotite micas, previously thought to have been the source of leached copper in porphyry copper deposits, are now shown be a sink. See page 516 and 472.

Double agent

Perforin, the molecule used by cytotoxic T cells to kill their targets, is shown to be homologous to a component of the serum complement cytolytic system, pages 525 and 475.

Guide to Authors

This issue, page 546.

ure (ISSN 0028-0836) is published weekly on Thursday, except the last week in December, Macmillan Magazines Ltd (4 Little Essex Street, London WC2R 3LF). Annual subscription USA and Canada US$250 (institutional/corporate), US$125 (individual making personal ment). USA and Canadian orders to: *Nature*, Subscription Dept, PO Box 7663, Teaneck, NJ 66-9857, USA. Other orders to *Nature*, Brunel Road, Basingstoke, Hants RG21 2XS, UK. ond class postage paid at New York, NY 10012 and additional mailing offices. Authorization photocopy material for internal or personal use, or internal or personal use of specific clients, ranted by *Nature* to libraries and others registered with the Copyright Clearance Center CC) Transactional Reporting Service, provided the base fee of $1.00 a copy plus $0.10 a page aid direct to CCC, 21 Congress Street, Salem, MA 01970, USA. Identification code for e: 0028-0836/88 $1.00 + $0.10. US Postmaster send address changes to: *Nature*, 65 cker Street, New York, NY 10012. Published in Japan by Nature Japan K.K., Shin-Mitsuke 36 Ichigaya Tamachi, Shinjuku-ku, Tokyo 162, Japan.

OPINION & NEWS

CORRESPONDENCE

NEWS AND VIEWS

SCIENTIFIC CORRESPONDENCE

BOOK REVIEWS

COMMENTARY

Contents continued ▶

lated, by defects in the processing machinery[7] or by the delivery of inhibitory signals[8,9]. *In vivo*, the antigen presenting function of T cells will generally be insignificant because very few proteins will bind directly to T cells. However, there are instances, for example the case of gp120, in which the situation might change dramatically. The fact that gp120 can bind to $CD4^+$ cells and be selectively presented could therefore have immunopathological consequences for HIV-1 infection. Because gp120 is readily shed from the surface of HIV-1 infected cells[10,11], the possibility exists that free gp120 might bind to uninfected $CD4^+$ T cells and macrophages and target them for destruction by gp120-specific cells. We are currently testing this possibility.

We thank Janette Millar for the preparation of the manuscript, and Drs Polly Matzinger, Ronald Palacios, Michael Bzay and Uwe Staerz for critical reading and comments. The Basel Institute for Immunology was founded and is supported by F. Hoffmann-La Roche & Co., Basel, Switzerland.

Received 11 May; accepted 24 June 1988.

1. Rock, K. L., Benacerraf, B. & Abbas, A. K. *J. exp. Med.* 160, 1102-1113 (1984).
2. Lanzavecchia, A. *Nature* 314, 537-539 (1985).
3. Lanzavecchia, A. *et al. J. exp. Med.* 167, 345-352 (1988).
4. Laski, L. A. *et al. Cell* 50, 975-985 (1987).
5. Klatzmann, D. *et al. Nature* 312, 767-768 (1984).
6. Sattentau, Q. J., Dalgleish, A. G., Weiss, R. A. & Beverly, P. C. L. *Science* 234, 1120-1123 (1986).
7. Gerrard, T. L., Volkman, D. J., Jurgensen, C. H. & Fauci, A. S. *Hum. Immun.* 17, 416-425 (1986).
8. Lamb, J. R. & Feldman, M. *Nature* 308, 72-74 (1984).
9. Rammensee, H. G., Bevan, M. J. & Fink, P. *Immun. Today* 6, 41-43 (1985).
10. Gelderbloom, H. R., Reupke, H. & Pauli, G. *Lancet* ii, 1016-1017 (1985).
11. Schneider, J., Kaaden, O., Copeland, T. D., Oroslan, S. & Hunsmann, G. *J. gen. Virol.* 67, 2533-2539 (1986).
12. Lasky, L. *et al. Science* 233, 209-212 (1986).
13. Roosnek, E., Demotz, S., Corradin, G. P. & Lanzavecchia, A. *J. Immun.* 140, 4079-4082 (1988).
14. Rabinovitch, P. S., June, C. H., Grossmann, A. & Ledbetter, J. A. *J. Immun.* 137, 952-961 (1986).

Identification of a protein encoded by the *vpu* gene of HIV-1

Eric A. Cohen*, Ernest F. Terwilliger*, Joseph G. Sodroski* & William A. Haseltine*†

* Dana-Farber Cancer Institute, Department of Pathology, Harvard Medical School, and † Harvard School of Public Health, 44 Binney Street, Boston, Massachusetts 02115, USA

Human immunodeficiency virus 1 (HIV-1) is the aetiological agent of AIDS[1-3]. The virus establishes lytic, latent and non-cytopathic productive infection in cells in culture[4,5]. The complexity of virus–host cell interaction is reflected in the complex organization of the viral genome[6-9]. In addition to the genes that encode the virion capsid and envelope proteins and the enzymes required for proviral synthesis and integration common to all retroviruses, HIV-1 is known to encode at least four additional proteins that regulate virus replication, the *tat*, *art*, *sor* and 3′ *orf* proteins, as well as a protein of unknown function from the open reading frame called R[10-18]. Close examination of the nucleic acid sequences of the genomes of multiple HIV isolates raised the possibility that the virus encodes a previously undetected additional protein. Here we report that HIV-1 encodes a ninth protein and that antibodies to this protein are detected in the sera of people infected with HIV-1. This protein distinguishes HIV-1 isolates from the other human and simian immunodeficiency viruses (HIV-2 and SIV)[19-21] that do not have the capacity to encode a similar protein.

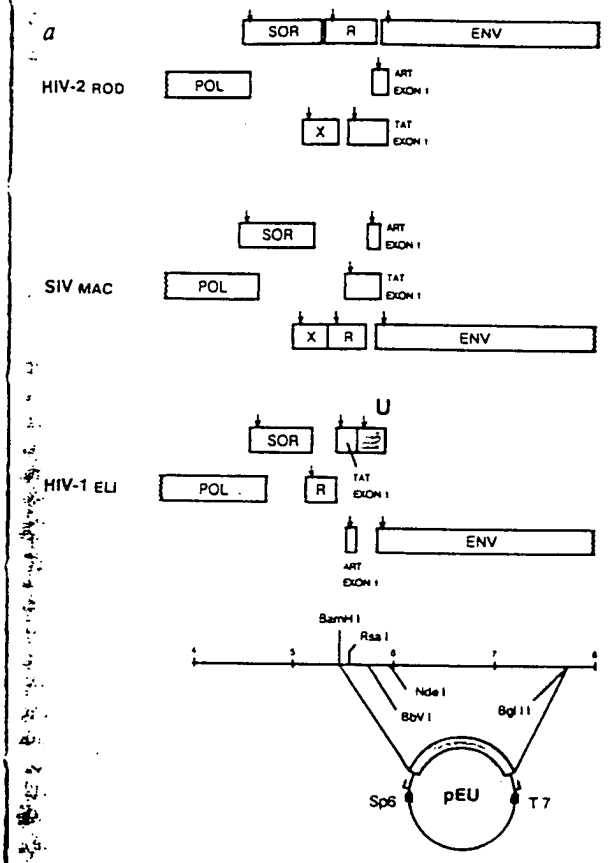
Figure 1*a* is a schematic diagram of the open reading frames of the region between the first coding exons of the *tat* and *art* genes of HIV-1 and the envelope glycoprotein gene. In this region many strains of the virus have the capacity to encode a protein of 80-82 amino acids that initiates with an AUG codon (Fig. 1*b*). To examine this possibility, two oligopeptides were made that correspond in sequence to regions of the protein which were predicted, on the basis of amino acid sequence, to be hydrophilic. One corresponded to amino acids 29 to 41 (peptide 1), and the other to amino acids 73 to 81 (peptide 2) (Fig. 1*b*). The amino acid sequences corresponded to the protein that BH10 substrain of the IIIB isolate was predicted to make[6]. The peptides were conjugated to keyhole limpet haemocyanin and used to raise antibody in three rabbits each. After multiple injections of the antigen, the rabbits were shown to produce antibodies that recognized the oligopeptide (data not shown).

The ability of the region between the first coding exon of *tat* and the *env* gene to encode a protein was first examined by an *in vitro* translation assay in a reticulocyte lysate[22], using RNA made *in vitro*[23]. RNA was made from a restriction fragment, 2,231 nucleotides long, of an HIV provirus that spanned the region between the first coding exons of the *tat*, *art* and part of the *env* genes. The template was derived from a fragment of the provirus of the ELI strain of HIV-1 placed 3′ to the SP6 bacteriophage RNA polymerase promoter[23] (Fig. 1*a*). This strain was selected as it contains an open reading frame in this region that initiates with an AUG codon (Fig. 1*b*)[24]. The viral sequences present in this RNA transcript, as shown in Fig. 1*a*, extend from the 5′ end of the first coding exon of the *tat* (*Bam*HI site) to 1,839 nucleotides (*Bgl*II site) within the *env* sequence. However, the initiation codon for the *tat* gene is not intact in this RNA as *Bam*HI cleaves the ELI proviral strain between the T and G residues of the *tat* initiation codon.

Proteins produced in the *in vitro* lysate using the RNA derived from this proviral fragment were labelled with ^{35}S-methionine and separated by size using sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) (Fig. 2*a*). The proteins synthesized in this sytem are displayed in lane 2. The proteins precipitated by rabbit anti-peptide-2 serum are also shown. Two proteins of relative molecular mass of approximately 15,000 (15K) and 16,000 (16K) are evident in the unfractionated extract and are precipitated by the rabbit antisera. The 15K and 16K proteins are not precipitated by the pre-immune rabbit sera (lane 3). All three of the antisera to peptide 2 recognize both proteins (lane 4) as do the antisera to peptide 1, albeit more weakly (data not shown). The data of Fig. 2*a* also show that peptide 2 competes for recognition of the 15K and 16K proteins by antisera (lane 5). However, peptide 1 (lane 6) or an unrelated peptide do not compete with anti-peptide-2 serum (lane 7).

To confirm the origin of the proteins, RNA from other proviral fragments was used in the *in vitro* translation assay. In one set of experiments, the template used for synthesis of RNA was truncated by restriction enzyme cleavage either seven nucleotides 5′ to the proposed AUG codon (*Rsa*I site) or 30 nucleotides 3′ to the proposed AUG codon (*Bbv*I site) (Fig. 1*a*). No specific protein products recognized by anti-peptide-2 antiserum were observed in these experiments (Fig. 2*b*, lanes 1 and 2). When the template used for synthesis of RNA was cleaved 102 nucleotides 3′ to the proposed stop codon (*Nde*I site), the 15K and 16K proteins were detected using anti-peptide-2 serum (Fig. 2*b*, lanes 3 and 4).

To examine the possibility that the proteins corresponding to the 15K and 16K products are produced in natural infections, the ability of antisera from normal and AIDS patients to recognize the protein synthesized *in vitro* was tested. The data of Fig. 2*c* demonstrate that HIV seropositive patient antisera recognize both the 15K and 16K proteins (lanes 2, 4 and 5). The ability of antiserum to precipitate the two proteins is partially competed out by peptide 2 (lane 3). The 15K and 16K proteins are not recognized by normal human serum (lane 1). However, all of the 19 sera of HIV-1 infected patients that immunoprecipitated the truncated *env* product were found to precipitate both the



b

PEPTIDE 1

```
                1         10 11        20 21        30 31        40 41        50
U ELI   MQPLGIIAIA ALVVAIILAI VVWTIVFIEY RRIKKQRRID CLLQRITERA
UHXB2   T**IP*V***V *******I** ***S**I*** *K*LR**K** R*I**LI***
U8H10   M**IQ*-***V *******I** ***K**I*** *K*LR**K** R*I**LI***
U8H8    M**IP*VT*V **A****I** ***K**I*** *K*LR**K** R*I**LI***
UHXB3   M**IQ*-***V *******I** ***K**I*** *K*LR**K** R*I**LI***
UMAL    I***V*L**V ****TL*I** *********I *K*RR**K** R*I**IR***
UBRU    M**IQ*-*** *******I** ***K**I*** *K*LR**K** R*I**LI***
USF2    M*S*Q*L**V S***VA*I** ******L*** *K*LR**K[I]* R*I***R*K*
```

PEPTIDE 2

```
                51        60 61        70 71        80 81 85
U ELI   EDSGNESEGD REKLSKL--- -VEMGHHAPW DIDDL  .   (81)
UHXB2   *********E ---I*A*VEM G********* *V***  .   (82)
U8H10   *********E ---I*A*VEM G********* *V***  .   (81)
U8H8    *********E ---I*A*VEM G********* *V***  .   (82)
UHXB3   *********E ---I*A*VEM G********* *V***  .   (81)
UMAL    ********** TEE*****--- -*****Q*** *V***  .   (81)
UBRU    *********E ---I*A*VEM G***-***** *I***  .   (80)
USF2    ********** QEEK*A*VEM G---**L*** *V***  .   (82)
```

Fig. 1 *a*, Genetic organization of the central region of HIV-1 (ELI isolate, ref. 24) compared with SIV[20,21] and HIV-2 (ROD isolate, ref. 19). Arrows indicate the initiator AUG codons in viral genes. SP6 plasmid used to synthesize messenger RNA, a *Bam*HI to *Bgl*II fragment, 2,231 nucleotides long, from the HIV ELI provirus that spanned the region between the first coding exons of the *tat*, *art* and part of the *env* gene was cloned 3′ to the SP6 bacteriophage RNA polymerase promoter[23]. Internal restriction sites used to linearize the plasmids are indicated. *b*, Aligment of the *vpu* gene protein sequence. The ELI isolate is taken as reference. Gaps (—) were introduced to optimize the alignment. Asterisks indicate amino acid identity. The HIV isolates compared include ELI, MAL (ref. 24), HXBc2, BH-10, BH-8, pHXB3 (ref. 6), BRU (ref. 7) and USF2 (ref. 8). USF2 contains a termination codon at position 39 ([I]). However, a −1 frameshift results in an extension of 43 amino acids that are well conserved when compared with the ELI U sequence.

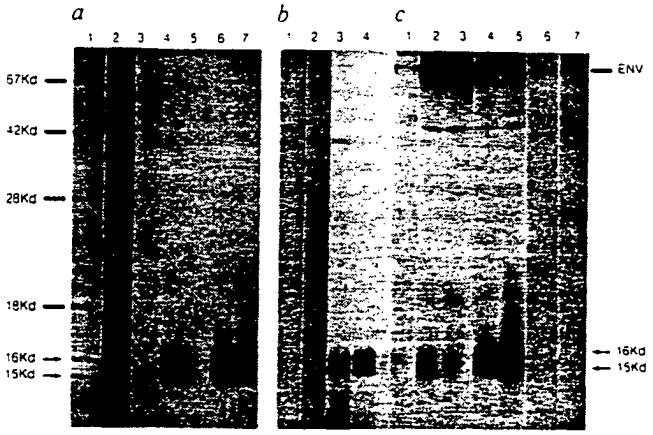


Fig. 2 *In vitro* characterization of the *vpu* gene product. *a*, pEU plasmid was linearized by digestion at an *Eco*RI site located in the polylinker 3′ to the HIV_{Eli} insert and used as template for *in vitro* transcription by SP6 RNA polymerase as described[26] except that the concentration of GTP and cap analogue m^7GpppG were raised to 0.2 and 1.0 mM respectively. Messenger RNAs were labelled with [5′-^{3}H) CTP and purified as described[26]. *In vitro* translation of equimolar amounts of RNA (equal amounts of radioactivity) was performed in reticulocyte lysate[22]. Incubation was at 30 °C for 30 min in the presence of ^{35}S-methionine. Labelled products were analysed directly by 15% SDS-PAGE (lane 2) or immunoprecipitated[27] beforehand with pre-immune rabbit serum (lane 3); anti-peptide-2 serum (lane 4); anti-peptide-2 serum in the presence of 500 μM of peptide 2 (lane 5), peptide 1 (lane 6), or an unrelated peptide QEEAETATKTSSC (lane 7). Lane 1 represent a total translation reaction with no mRNA added. *b*, pEU plasmid was linearized with the following restriction enzymes *Rsa*I (lane 1); *Bbv*I (lane 2) and *Nde*I (lanes 3 and 4). SP6-generated RNAs were translated *in vitro* and immunoprecipitation was performed on the labelled products using anti-peptide-2 serum (lanes 1, 2 and 4). Lane 3 represents a total *in vitro* translation reaction. *c*, After *in vitro* translation of SP6-generated pEU RNA, the labelled products were immunoprecipitated as described[27] except that 1 M NaCl was used in the immunoprecipitation reaction. Immunoprecipitation with a pool of normal human serum (lane 1); HIV-1-infected human sera (lanes 2, 4 and 5); HIV-1-infected patient serum in the presence of 500 μM of peptide 2 (lane 3); HIV-2-infected human serum (lane 6) or SIV-infected *Rhesus macaques* serum (lane 7). Fifteen HIV-2-infected human serum and four SIV-infected *Rhesus macaques* serum were tested. These sera were demonstrated to specifically react with HIV-2 or SIV proteins by immunoprecipitation and Western blot analysis (not shown). None of these antisera immunoprecipitated $p15^{vpu}$ and $p16^{vpu}$. Immunoprecipitates were resolved on 15% SDS-PAGE.

15K and 16K proteins (data not shown). Antisera from HIV-2-infected humans or from SIV-infected macaques do not precipitate either protein (Fig. 2*c*, lanes 6 and 7).

We examined whether the anti-peptide-2 serum recognized the 15K and 16K proteins in three cell lines that constitutively express HIV-1 proteins *art* and *env* encoded by the 3′ half of the virus. Cloned HeLa cell lines that have stably integrated the region between the *art* gene and the 3′ long terminal repeat (LTR) of the proviral ELI, HXBc2 and MAL strains[24,6] of HIV were isolated (Terwilliger *et al*, in preparation). The plasmids used for construction of these cell lines contained the HIV LTR juxtaposed 5′ to the initiation codon of the *art* gene. The *tat* gene product was supplied in *trans*. Figure 3*b* shows that the anti-peptide-2 antiserum specifically recognized a 15K protein in the cell line derived from the ELI provirus (lane 3) that comigrates with the 15K protein made *in vitro* (lane 2). The same antiserum does not recognize a protein in the cell line that expresses proteins derived from the MAL (Fig. 3*c*) or the HXBc2

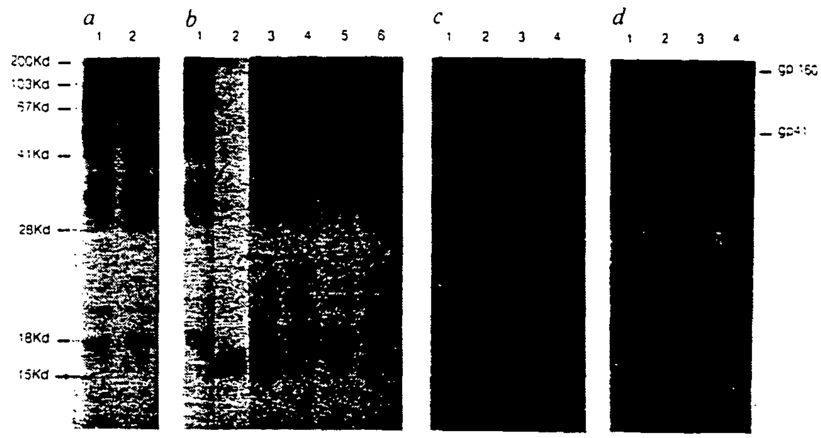


Fig. 3 Identification of the *vpu* gene product in cell lines constitutively expressing proteins encoded by the 3' half of the HIV-1 ELI provirus. HeLa cell lines that have the region of the HIV-1 provirus between the *art* gene and the 3' LTR of the HIV proviral ELI, HXBc2 and MAL strains stably integrated were constructed (Terwilliger *et al.*, in preparation). The parental cells used to isolate these cell lines had previously been selected for constitutive expression of the *tat* gene product, following infection with a retroviral vector carrying the *tat* coding sequences[28]. Cells were labelled with ^{35}S-methionine and cysteine and cell lysates were immunoprecipitated as described[27]. *a*, HeLa *tat* cell line lysates immunoprecipitated with anti-peptide-2 serum (lane 1) or HIV-1-infected patient serum (lane 2). *b*, HeLa *tat* ELI lysates immunoprecipitated with pre-immune rabbit serum (lane 1); antipeptide-2 serum (lane 3); anti-peptide-2 serum in the presence of 500 μM of peptide 2 (lane 4); normal human serum (lane 5); HIV-1-infected patient serum (lane 6). Lane 2 represent an immunoprecipitation of labelled *in vitro* translated product from pEU RNA with anti-peptide-2 serum. *c*, HeLa *tat* MAL lysate immunoprecipitated with pre-immune rabbit serum (lane 1); anti-peptide-2 serum (lane 2): normal human serum (lane 3) and HIV-1-infected patient serum (lane 4). *d*, HeLa *tat* IIIB lysate immunoprecipitated with pre-immune rabbit serum (lane 1); anti-peptide-2 serum (lane 2) normal human serum (lane 3) and HIV-1-infected patient serum (lane 4).

(Fig. 3*d*) proviruses. This result was expected as neither of the proviruses contain a properly positioned initiation codon at the 5' end of the open reading frame (Fig. 1*b*). The absence of detection of the 15K protein by the HIV-1 patient antiserum in the cell line derived from the ELI provirus is probably due to both the low antibody titre in the antiserum used and the much smaller amount of the 15K protein in the cell line compared to the *in vitro* translation products.

The experiments presented here demonstrate that HIV-1 has the capacity to encode a previously unrecognized protein. The open reading frame from which this protein is synthesized was originally designated U (ref. 7) and so we propose to call the gene *vpu*, for viral protein U, and the proteins produced $p15^{vpu}$ and $p16^{vpu}$. The product of *vpu* is made upon HIV-1 infection as antisera from the majority of HIV-1-infected people surveyed have antibodies that recognize the protein.

All HIV-1 proviral strains isolated contain an open reading frame in the region corresponding to *vpu*. However, the ability of the individual proviral strains to produce a protein from this region is compromised in some strains by a single point mutation that prevents *vpu* expression. Indeed, different proviral strains from the same viral isolate differ in their ability to encode *vpu*: independent proviral clones of the IIIB isolate, HXBc2, BH10, BH-8 and BH-3 are an example (Fig. 1*b*). There is a similar variation in the ability of individual proviral clones to encode other viral proteins, for example, the 3' *orf* product. The mutation that truncates the protein product of the IIIB 3' *orf* yields a virus that replicates more rapidly in culture than the wild-type virus[15]. The virus produced by transfection with HXBc2 can grow in T cells in culture[25] implying that a virus which cannot express *vpu* can replicate. However, the ability of the vpu^- virus to replicate does not rule out the possibility that the *vpu* product is important in regulation of viral replication or pathogenesis.

The *vpu* gene distinguishes HIV-1 from HIV-2 and SIV infections. A computer-assisted search for proteins similar to $p15/16^{vpu}$ showed that HIV-2 and SIV do not encode a similar protein. HIV-2 and SIV strains do contain an open reading frame that is missing from that of HIV-1 isolates, the X open reading frame[19], but there is no predictable similarity in the predicted protein products of *vpu* and the X open reading frame. None of the sera of HIV-2-infected patients surveyed contained antibodies to the *vpu* product, nor were antibodies to *vpu* detected in *Rhesus macaques* infected with SIV.

We note that *vpu* is highly conserved amongst the HIV-1 proviral sequences isolated (Fig. 1*b*), and that *vpu* is removed by splicing from viral messenger RNAs that encode regulatory proteins[9,11,16]. It is therefore predicted that the *vpu* product is not made in the absence of the *art* gene product as only fully spliced messenger RNAs accumulate in the absence of this product[16,17]. We suspect that the *vpu* product is made late in infection like virion proteins.

We thank Yassamin Jalinoos and Jeff Proulx for technical assistance, Luc Montagnier for HIV ELI and MAL proviruses: Robert Schooley and Phyllis Kanki for antisera from HIV-1 and HIV-2-infected patients and simian antisera and Temple Smith for helpful discussions. This research was supported by a grant from the National Institutes of Health to W.A.H. E.A.C. is a recipient of a fellowship from National Cancer Institute of Canada. E.T. is a fellow of the Leukemia Society of America. J.S. is a scholar of the Leukemia Society of America and is supported by the American Foundation for AIDS Research and the National Institute of Health.

Received 27 April; accepted 5 July 1988.

1. Barre-Sinoussi, F. *et al. Science* **220**, 868-871 (1983).
2. Gallo, R. *et al. Science* **224**, 500-503 (1984).
3. Levy, J. A. *et al. Science* **225**, 840-842 (1984).
4. Curran, J. W. *et al. Science* **223**, 1352-1357 (1985).
5. Zagury, D. *et al. Science* **231**, 850-853 (1986).
6. Ratner, L. *et al. Nature* **313**, 277-283 (1985).
7. Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. & Alizon, M. *Cell* **40**, 9-17 (1985).
8. Sanchez-Pescador, R. *et al. Science* **227**, 484-492 (1985).
9. Muesing, M. A. *et al. Nature* **313**, 450-458 (1985).
10. Sodroski, J. G. *et al. Science* **227**, 171-173 (1985).
11. Arya, S. K., Guo, C., Josephs, S. F. & Wong-Staal, F. *Science* **229**, 69-73 (1985).
12. Lee, T. H. *et al. Science* **231**, 1546-1549 (1986).
13. Sodroski, J. *et al. Science* **231**, 1549-1551 (1986).
14. Allan, J. S. *et al. Science* **230**, 810-813 (1985).
15. Terwilliger, E., Sodroski, J. G., Rosen, C. A. & Haseltine, W. A. *J. Virol.* **60**, 754-760 (1986).
16. Sodroski, J. *et al. Nature* **321**, 413-417 (1986).
17. Feinberg, M. B., Jarrett, R. F., Aldovini, A., Gallo, R. C. & Wong-Staal, F. *Cell* **46**, 807-817 (1986).
18. Wong-Staal, F., Chanda, P. K. & Ghrayeb, J. *AIDS Res. hum. Retroviruses* **3**, 33-39 (1987).
19. Guyader, M. *et al. Nature* **326**, 662-669 (1987).
20. Hirsch, V., Riedel, N. & Mullins, J. I. *Cell* **49**, 307-319 (1987).
21. Chakrabarti, L. *et al. Nature* **328**, 543-547.
22. Pelham, H. P. B. and Jackson, R. S. *Eur. J. Biochem.* **67**, 247-256 (1986).
23. Melton, D. A. *et al. Nucleic Acids Res.* **12**, 7035-7056 (1984).
24. Alizon, M., Wain-Hobson, S., Montagnier, L. & Sonigo, P. *Cell* **46**, 63-74 (1986).
25. Fisher, A. G., Collati, E., Ratner, L., Gallo, R. C. & Wong-Staal, F. *Nature* **316**, 262-265 (1985).
26. Pelletier, J. & Sonnenberg, N. *Cell* **40**, 515-526 (1985).
27. Lee, T. H. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 7579-7583 (1984).
28. Rosen, C. A., Sodroski, J. G., Campbell, K. & Haseltine, W. A. *J. Virol.* **57**, 379-384 (1986).

Cell

Editorial Office
Cell
292 Main Street
Cambridge, Massachusetts 02142
USA
617-253-2890
Telex 314765

European Office
Cancer Research Campaign
Eukaryotic Molecular Genetics Group
Department of Biochemistry
Imperial College of Science
and Technology
London SW7 2AZ, England
01-584-9913

Cell is published monthly from January to November and twice monthly in December by The MIT Press, Cambridge, Massachusetts, and London, England. Subscriptions are available by the calendar year. The order rate direct from the publisher is $225 (USA and Canada) or $245 (elsewhere) for 1985 (volumes 40–43). Back issue rates for 1982–1984 are available on request. Subscription correspondence should be addressed to: The MIT Press Journals Department, 28 Carleton Street, Cambridge, Massachusetts 02142 (617-253-2889).

A charge of $35 per page is made for publication. Inability to pay will not influence decisions on acceptance, and authors unable to meet this charge should make the reason known upon publication. Second class postage paid at Boston, Massachusetts, and additional mailing offices. Postmaster: send address changes to Cell (ISSN 0092-8674), 28 Carleton Street, Cambridge, Massachusetts 02142.

Cell, Vol. 40, 9–17, January 1985, Copyright © 1985 by MIT 0092-8674/85/010009-09 $02.00/0

Nucleotide Sequence of the AIDS Virus, LAV

Simon Wain-Hobson,* Pierre Sonigo,*
Olivier Danos,† Stewart Cole,‡ and Marc Alizon§
*Unité de Recombinaison et Expression Génétique
†Unité des Virus Oncogènes
‡Groupement de Génie Génétique
§Unité d'Oncologie Virale
Institut Pasteur
25 et 28 rue du Dr. Roux
75724 Paris, Cedex 15, France

Summary

The complete 9193-nucleotide sequence of the probable causative agent of AIDS, lymphadenopathy-associated virus (LAV), has been determined. The deduced genetic structure is unique: it shows, in addition to the retroviral gag, pol, and env genes, two novel open reading frames we call Q and F. Remarkably, Q is located between pol and env and F is half-encoded by the U3 element of the LTR. These data place LAV apart from the previously characterized family of human T cell leukemia/lymphoma viruses.

Introduction

The recent onset of severe opportunistic infections among previously healthy male homosexuals has led to the characterization of the acquired immune deficiency syndrome (AIDS) (Gottlieb et al., 1981; Masur et al., 1981). The disease has spread dramatically, and new high-risk groups have been identified: patients receiving blood products, intravenous drug addicts, and individuals originating from Haiti and Central Africa (Piot et al., 1984). AIDS is a fatal disease, and there is at present no specific treatment. The causative agent was suspected to be of viral origin since the epidemiological pattern of AIDS was consistent with a transmissible disease, and cases had been reported after treatment involving ultrafiltered anti-hemophilia preparations (Daly and Scott, 1983). A decisive step in AIDS research was the discovery of a novel human retrovirus called lymphadenopathy-associated virus (LAV) (Barré-Sinoussi et al., 1983). The properties of the virus consistent with its etiological role in AIDS are: the recovery of many independent isolates from patients with AIDS or related diseases (Montagnier et al., 1984); high LAV seropositivity among these populations (Brun-Vézinet et al., 1984); a tropism and cytopathic effect in vitro for the helper/inducer T-lymphocyte subset T4 (Klatzmann et al., 1984), also found depleted in vivo.

Other groups have reported the isolation of human retroviruses, the human T cell leukemia/lymphoma/lymphotropic virus type III (HTLV-III) (Popovic et al., 1984) and the AIDS-associated retrovirus (ARV), which display biological and sero-epidemiological properties very similar to if not identical with those of LAV (Levy et al., 1984; Popovic et al., 1984; Schüpbach et al., 1984). Both LAV and HTLV-III genomes have been molecularly cloned (Alizon et al., 1984; Hahn et al., 1984). Their restriction maps show remarkable agreement, including a Hind III restriction site polymorphism, bearing in mind the variability of this virus (Shaw et al., 1984) and confirming that these two viruses represent a single viral lineage.

In addition to its obvious diagnostic and therapeutic potential, the LAV DNA nucleotide sequence is essential to an understanding of the genetics and molecular biology of the virus and its classification among retroviruses. We report here the complete 9193-nucleotide sequence of the LAV genome established from cloned proviral DNA.

Results

DNA Sequence and Organization of the LAV Genome

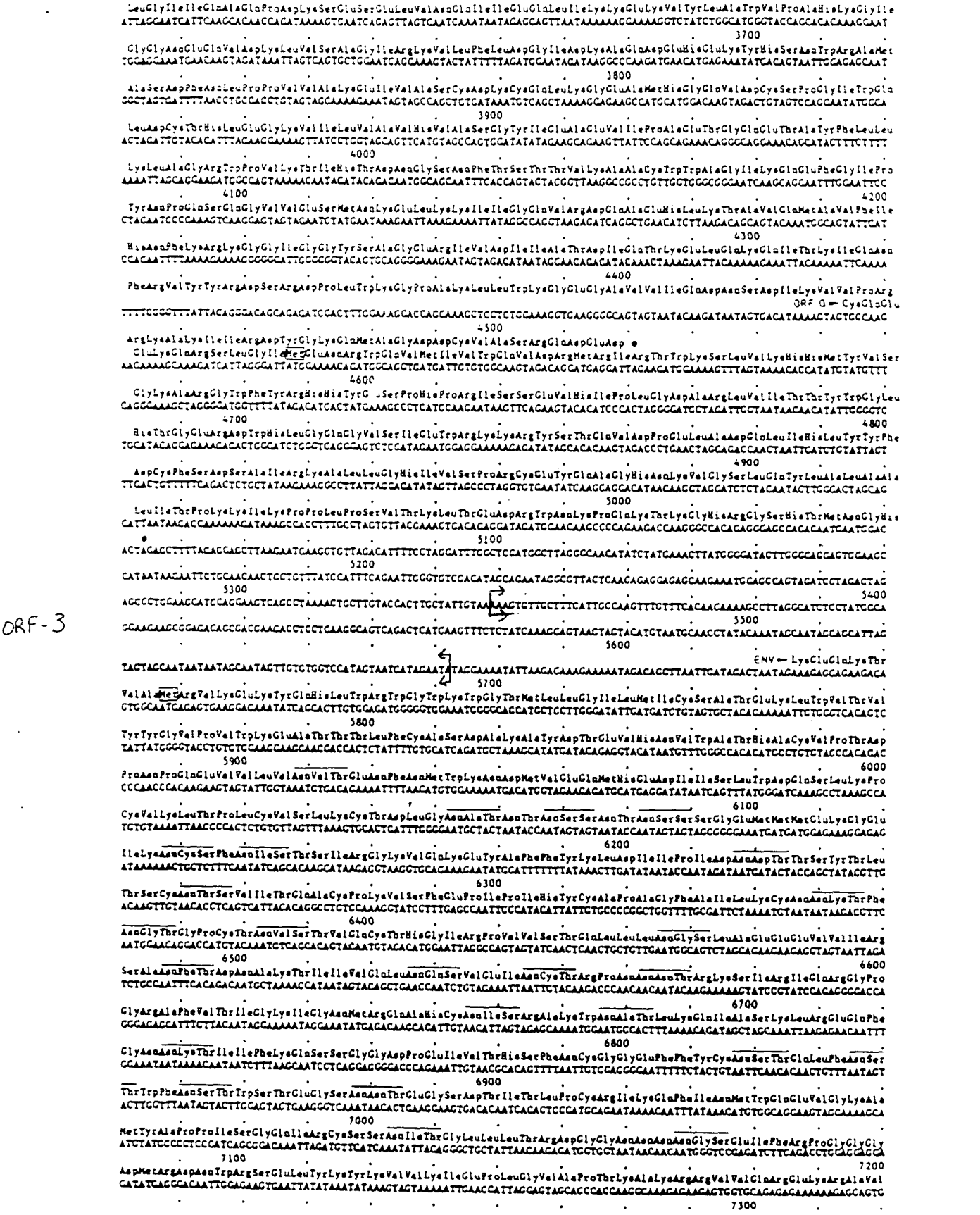
We have reported previously the molecular cloning of both cDNA and integrated proviral forms of LAV (Alizon et al., 1984). The recombinant phage clones were isolated from a genomic library of LAV-infected human T-lymphocyte DNA partially digested by Hind III. The insert of recombinant phage λJ19 was generated by Hind III cleavage within the R element of the long terminal repeat (LTR). Thus each extremity of the insert contains one part of the LTR. We have eliminated the possibility of clustered Hind III sites within R by sequencing part of an LAV cDNA clone, pLAV 75 (Alizon et al., 1984), corresponding to this region (data not shown). Thus the total sequence information of the LAV genome can be derived from the λJ19 clone.

Using the M13 shotgun cloning and dideoxy chain termination method (Sanger et al., 1977), we have determined the nucleotide sequence of λJ19 insert. The reconstructed viral genome with two copies of the R sequence is 9193 nucleotides long. The numbering system starts at the cap site (see below) of virion RNA (Figure 1).

The viral (+) strand contains the statutory retroviral genes encoding the core structural proteins (gag), reverse transcriptase (pol), and envelope protein (env), and two extra open reading frames (orf) that we call Q and F (Table 1). The genetic organization of LAV, 5'LTR-gag-pol-Q-env-F-3'LTR, is unique. Whereas in all replication-competent retroviruses pol and env genes overlap, in LAV they are separated by orf Q (192 amino acids) followed by four small (<100 triplets) orf. The orf F (206 amino acids) slightly overlaps the 3' end of env and is remarkable in that it is half-encoded by the U3 region of the LTR.

Such a structure clearly places LAV apart from previously sequenced retroviruses (Figure 2). The (−) strand is apparently noncoding. The additional Hind III site of the LAV clone λJ81 (with respect to λJ19) maps to the apparently noncoding region between Q and env (positions 5166–5745). Starting at position 5501 is a sequence (AAGCCT) that differs by a single base (underlined) from the Hind III recognition sequence. It is anticipated that many of the restriction site polymorphisms between different isolates will map to this region.


HindIII
100
200
GAG ← LeuAlaGluAlaArgArgArgLeuGlyAlaArgAlaSerValLeuSer
300
GlyGlyGluLeuAspArgTrpGluLysIleArgLeuArgProGlyGlyLysLysLysTyrLysLeuLysHisIleValTrpAlaSerArgGluLeuGluArgPheAlaValAsnProGly
400
LeuLeuGluThrSerGluGlyCysArgGlnIleLeuGlyGlnLeuGlnProSerLeuGlnThrGlySerGluGluLeuArgSerLeuTyrAsnThrValAlaThrLeuTyrCysValHis
500
600
GlnArgIleGluIleLysAspThrLysGluAlaLeuAspLysIleGluGluGluGlnAsnLysSerLysLysLysAlaGlnGlnAlaAlaAlaAspThrGlyHisSerSerGlnValSer
700
GlnAsnTyrProIleValGlnAsnIleGlnGlyGlnMetValHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysValValGluGluLysAlaPheSerProGluValIle
800
ProMetPheSerAlaLeuSerGluGlyAlaThrProGlnAspLeuAsnThrMetLeuAsnThrValGlyGlyHisGlnAlaAlaMetGlnMetLeuLysGluThrIleAsnGluGluAla
900
AlaGluTrpAspArgValHisProValHisAlaGlyProIleAlaProGlyGlnMetArgGluProArgGlySerAspIleAlaGlyThrThrSerThrLeuGlnGluGlnIleGlyTrp
1000
MetThrAsnAsnProProIleProValGlyGluIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSerProThrSerIleLeuAspIleArgGlnGlyPro
1100
1200
LysGluProPheArgAspTyrValAspArgPheTyrLysThrLeuArgAlaGluGlnAlaSerGlnGluValLysAsnTrpMetThrGluThrLeuLeuValGlnAsnAlaAsnProAsp
1300
CysLysThrIleLeuLysAlaLeuGlyProAlaAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProGlyHisLysAlaArgValLeuAlaGluAlaMetSerGln
1400
ValThrAsnSerAlaThrIleMetMetGlnArgGlyAsnPheArgAsnGlnArgLysIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArgAsnCysArgAlaProArg
1500
POL ← PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSer
LysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerTyrLysGlyArgProGlyAsnPheLeu
1600
SerGluGlnThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSerLeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPhePro
GlnSerArgProGluProThrAlaProProGluGluSerPheArgSerGlyValGluThrThrThrProSerGlnLysGlnGluProIleAspLysGluLeuTyrProLeuThrSerLeu
1700
1800
GlnIleThrLeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArg
ArgSerLeuPheGlyAsnAspProSerSerGln ●
1900
TrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrPro
2000
ValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLysValLys
2100
GlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAla
2200
IleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLys
2300
2400
LysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIle
2500
ArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGln
2600
TyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGln
2700
LysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGly
2800
LysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGlu
2900
3000
LeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyr
3100
GlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgThrArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIle
3200
TrpGlyLysThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeu
3300
ValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAlaSerArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGly
3400
ArgGlnLysValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAla
3500
3600


ORF-3

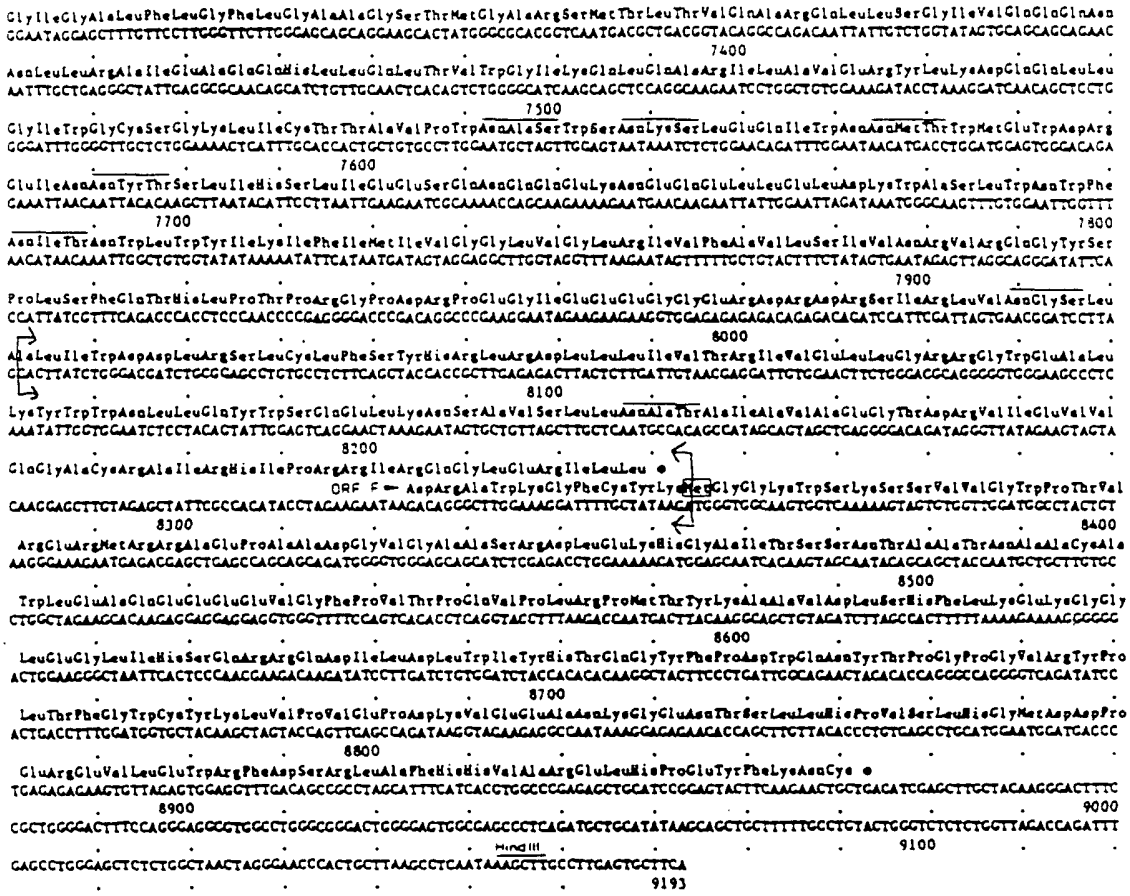


Figure 1. Complete DNA Sequence of Viral Genome (LAV-1a)

The sequence was reconstructed from the sequence of phage λJ19 insert. The numbering starts at the cap site, which was located experimentally (see above). Important genetic elements, major open reading frames, and their predicted products are indicated together with the Hind III cloning sites. The potential glycosylation sites in the env gene are overlined. The NH_2- terminal sequence of $p25^{gag}$ determined by protein microsequencing is boxed (Genetic Systems, personal communication).

Each nucleotide was sequenced on average 5.3 times: 85% of the sequence was determined on both strands and the remainder was sequenced at least twice from independent clones. The base composition is T, 22.2%; C, 17.8%; A, 35.8%; G, 24.2%; G + C, 42%. The dinucleotide CpG is greatly under-represented (0.9%) as is common among eukaryotic sequences (Bird, 1980).

The LTR

The organization of a reconstructed LTR and viral flanking elements are shown schematically in Figure 3. The LTR is 638 bp long and displays usual features (Chen and Barker, 1984): it is bounded by an inverted repeat (5'ACTG) including the conserved TG dinucleotide (Temin, 1981); adjacent to 5' LTR is the tRNA primer binding site (PBS), complementary to $tRNA^{Lys}_3$ (Raba et al., 1979); adjacent to 3' LTR is a perfect 15 bp polypurine tract. The other three polypurine tracts observed between nucleotides 8200–8800 are not followed by a sequence that is complementary to that just preceding the PBS.

The limits of U5, R, and U3 elements were determined as follows. U5 is located between PBS and the polyadenylation site established from the sequence of the 3' end of oligo(dT)-primed LAV cDNA (Alizon et al., 1984). Thus U5 is 84 bp long. The length of R+U5 was determined by synthesizing tRNA-primed LAV cDNA. After alkaline hydroly-

Table 1. Locations and Sizes of Viral Open Reading Frames

orf	1st Triplet	Met	Stop	No. Amino Acids	M, Calc.
gag	312	336	1,836	500	55,841
pol	1,631	1,934	4,640	(1,003)	(113,629)
orf Q	4,554	4,587	5,163	192	22,487
env	5,746	5,767	8,350	861	97,376
orf F	8,324	8,354	8,972	206	23,316

The nucleotide coordinates refer to the first base of the first triplet (1st triplet), of the first methionine (initiation) codon (Met) and of the stop codon (Stop). The numbers of amino acids and molecular weights are those calculated for unmodified precursor products starting at the first methionine through to the end, with the exception of pol, where the size and M, refer to that of the whole orf.

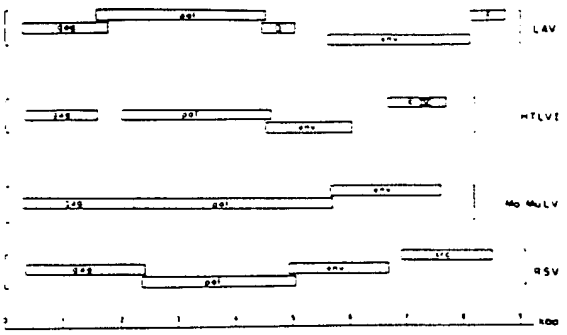


Figure 2. Comparison of the Genome Organization of LAV with Those of Human T Cell Leukemia/Lymphoma Virus Type I (HTLV-I) (Seiki et al., 1983), Moloney Murine Leukemia Virus (MoMuLV) (Shinnick et al., 1981), and Rous Sarcoma Virus (RSV) (Schwartz et al., 1983)

The positions and sizes of viral genes are drawn to scale (open boxes) and the viral genomes (RNA forms) are delimited by brackets.

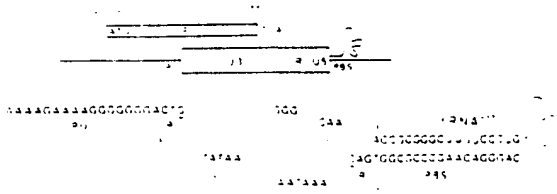


Figure 3. Schematic Representation of the LAV Long Terminal Repeat (LTR)

The LTR was reconstructed from the sequence of LJ19 by juxtaposing the sequences adjacent to the Hind III cloning sites. Sequencing of oligo(dT)-primed LAV DNA clone pLAV75 (Alizon et al., 1984) rules out the possibility of clustered Hind III sites in the R region of LAV. LTR are limited by an inverted repeat sequence (IR). Both of the viral elements flanking the LTR have been represented as tRNA primer binding site (PBS) for 5' LTR and polypurine track (PU) for 3' LTR. Also indicated are a putative TATA box, the cap site, polyadenylation signal (AATAAA), and polyadenylation site (CAA). The location of the open reading frame F (648 nucleotides) is shown above the LTR scheme.

sis of the primer, R+U5 was found to be 181 ±1 bp (Figure 4). Thus R is 97 bp long and the cap site at its 5' end can be located. Finally, U3 is 456 bp long. The LAV LTR also contains characteristic regulatory elements: a polyadenylation signal sequence AATAAA 19 bp from the R-U5 junction, and the sequence ATATAAG, which is very likely the TATA box, 22 bp 5' of the cap site. There are no long direct repeats within the LTR. Interestingly, the LAV LTR shows some similarities to that of the mouse mammary tumor virus (MMTV) (Donehower et al., 1981). They both use $tRNA^{lys}_3$ as a primer for (−) strand synthesis, whereas all other exogenous mammalian retroviruses known to date use $tRNA^{pro}$ (Chen and Barker, 1984). They possess very similar polypurine tracts; that of LAV is AAAAGAAAAGGGGGG while that of MMTV is AAAAAAGAAAAAAGGGGG. It is probable that the viral (+) strand synthesis is discontinuous since the polypurine tract flanking the U3 element of the 3'LTR is found exactly duplicated in the 3' end of orf pol, at 4331–4346. In addition, MMTV and LAV are exceptional in that the U3 element can encode an orf. In the case of MMTV, U3 contains the whole orf while, in LAV, U3 contains 110 codons of the 3' half of orf F.

Viral Proteins

gag

Near the 5' extremity of the gag orf is a "typical" initiation codon (Kozak, 1984) (position 336), which is not only the first in the gag orf, but the first from the cap site. The precursor protein is 500 amino acids long. The calculated M_r of 55,841 agrees with the 55 kd gag precursor polypeptide (Luc Montagnier, unpublished results). The N-terminal amino acid sequence of the major core protein p25, obtained by microsequencing (Genetic Systems, personal communication), matches perfectly with the translated nucleotide sequence starting from position 732 (see Figure 1). This formally makes the link between the cloned LAV genome and the immunologically characterized LAV p25 protein. The protein encoded 5' of the p25 coding sequence is rather hydrophilic. Its calculated M_r of 14,866 is consistent with that of the gag protein p18. The 3' part of the gag region probably codes for the retroviral nucleic acid binding protein (NBP). Indeed, as in HTLV-I (Seiki et al., 1983) and RSV (Schwartz et al., 1983), the motif Cys-X_2-Cys-X_{4-5}-Cys common to all NBP (Oroszlan et al., 1984) is found duplicated (nucleotides 1509 and 1572 in LAV sequence). Consistent with its function the putative NBP is extremely basic (17% Arg + Lys).

pol

The reverse transcriptase gene can encode a protein of up to 1003 amino acids (calculated M_r = 113,629). Since the first methionine codon is 92 triplets from the origin of the open reading frame, it is possible that the protein is translated from a spliced messenger RNA, giving a gag-pol polyprotein precursor.

The pol coding region is the only one in which significant homology has been found with other retroviral protein sequences, three domains of homology being apparent. The first is a very short region of 17 amino acids (starting at 1856). Homologous regions are located within the p15 gag^{RSV} protease (Dittmar and Moelling, 1978) and a polypeptide encoded by an open reading frame located between gag and pol of HTLV-I (Figure 5) (Schwartz et al., 1983; Seiki et al., 1983). This first domain could thus correspond to a conserved sequence in viral proteases. Its different locations within the three genomes may not be significant since retroviruses, by splicing or other mechanisms, express a gag-pol polyprotein precursor (Schwartz et al., 1983; Seiki et al., 1983). The second and most extensive region of homology (starting at 2048) probably represents the core sequence of the reverse transcriptase. Over a region of 250 amino acids, with only minimal insertions or deletions, LAV shows 38% amino acid identity with RSV, 25% with HTLV-I, and 21% with MoMuLV (Schinnick et al., 1981) while HTLV-I and RSV show 38% identity in the same region. A third homologous region is situated at the 3' end of the pol reading frame and corresponds to part of the pp32 peptide of RSV that has exonuclease activity (Misra et al., 1982). Once again, there is greater homology with the corresponding RSV sequence than with HTLV-I.

env

The env open reading frame has a possible initiator methionine codon very near the beginning (eighth triplet).

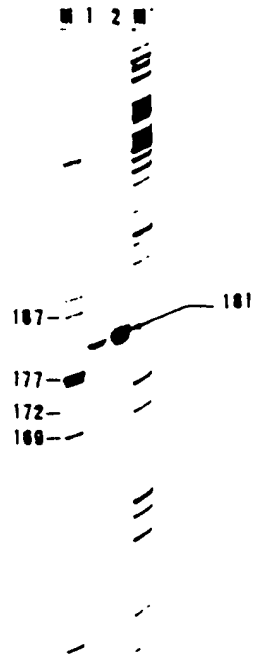


Figure 4. Synthesis of RNA-Primed LAV cDNA for R+U5 (Strong-Stop cDNA)

Lanes 1 and 2 show two different quantities of cDNA while lanes M and M' represent markers. The strong-stop cDNA is 181 bases long with a second, less intense band at 180. The error of estimation is ±1 bp. This maps the major cap site to the second G residue of the sequence CTGGGTCT within the LTR, 24 nucleotides downstream of the TATA box. This guanosine residue is taken as the first base in the nucleotide sequence shown in Figure 1.

If so, the molecular weight of the presumed env precursor protein (861 amino acids, M_r calc = 97,376) is consistent with the known size of the LAV glycoprotein (110 kd and 90 kd after glycosidase treatment; Luc Montagnier, unpublished). There are 32 potential N-glycosylation sites (Asn-X-Ser/Thr), which are overlined in Figure 1. An interesting feature of env is the very high number of Trp residues at both ends of the protein. There are three hydrophobic regions, characteristic of the retroviral envelope proteins (Seiki et al., 1983), corresponding to a signal peptide (encoded by nucleotides 5815–5850 bp), a second region (7315–7350 bp), and a transmembrane segment (7831–7896 bp). The second hydrophobic region (7315–7350 bp) is preceded by a stretch rich in Arg + Lys. It is possible that this represents a site of proteolytic cleavage, which, by analogy with other retroviral proteins, would give an external envelope polypeptide and a membrane-associated protein (Seiki et al., 1983; Kiyokawa et al., 1984). A striking feature of the LAV envelope protein sequence is that the region following the transmembrane segment is of unusual length (150 residues). The env protein shows no

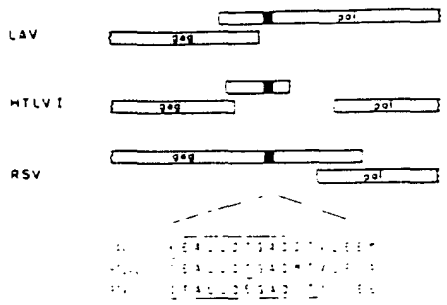


Figure 5. Location of a Short Stretch of Homology in the gag-pol Region of the LAV, HTLV-I (Seiki et al., 1983) and RSV (Schwartz et al., 1983) Genomes

Conserved amino acids are boxed. Homologous region is shown by the solid bar in the schema. Each virus is organized differently in this region but the sequence in the RSV genome maps to $p15^{gag}$, which has a protease-associated function.

homology to any sequence in protein data banks. The small amino acid motif common to the transmembrane proteins of all leukemogenic retroviruses (Cianciolo et al., 1984) is not present in LAV env.

Q and F

The location of orf Q is without precedent in the structure of retroviruses. Orf F is unique in that it is half-encoded by the U3 element of the LTR. Both orf have strong initiator codons (Kozak, 1984) near their 5' ends and can encode proteins of 192 amino acids (M_r calc = 22,487) and 206 amino acids (M_r calc = 23,316), respectively. Both putative proteins are hydrophilic (pQ 49% polar, 15.1% Arg + Lys; pF 46% polar, 11% Arg + Lys) and are therefore unlikely to be associated directly with membrane. The function for the putative proteins pQ and pF cannot be predicted, as no homology was found by screening protein sequence data banks. Between orf F and the pX protein of HTLV-I there is no detectable homology. Furthermore, their hydrophobicity/hydrophilicity profiles are completely different. It is known that retroviruses can transduce cellular genes—notably proto-oncogenes (Weinberg, 1982). We suggest that orfs Q and F represent exogenous genetic material and not some vestige of cellular DNA because LAV DNA does not hybridize to the human genome under stringent conditions (Alizon et al., 1984), and their codon usage is comparable to that of the gag, pol, and env genes (data not shown).

Relationship to Other Retroviruses

Although LAV is both morphologically and biochemically (Barré-Sinoussi et al., 1983) distinct to HTLV-I and -II, it remained possible that its genome was organized in a similar manner. The characteristic features of HTLV-I and -II genomes, which they share with the more distantly related bovine leukemia virus (BLV) (Rice et al., 1984), are not observed in the case of LAV. These are: a region 3' of the envelope gene consisting of a noncoding stretch (600–900 bp), followed by a coding sequence of 307–357 codons (X open reading frame), which may slightly overlap the U3 region of the LTR (Seiki et al., 1983; Rice et al., 1984; Sagata et al., 1984) and, second, the LTR being

Table 2. Comparison of the Size of the LAV LTR and LTR-Related Element to Those of Other Retroviruses

	LTR	U3	R	U5	PU	PBS	iR
LAV	638	456	97	85	15	LYS	4
HTLV-I	759	355	228	176	12'	PRO	4'
HTLV-II	763	314	248	261	12'	PRO	4'
MMTV	1.332	1.197	11	124	19	LYS	8'
MoMuLV	594	449	68	77	13	PRO	13
RSV	335	234	21	80	11	TRP	15
SNV	601	420	97	80	13	PRO	9

Adapted from Chen and Barker (1984).
i = imperfect match or tract.
SNV = spleen necrosis virus (Shimotohno and Temin, 1982).

composed of unusually long U5 and R elements and the polyadenylation signal being situated in U3 instead of R (Seiki et al., 1983; Sagata et al., 1984; Shimotohono et al., 1984). We show here that, in contrast, the 3' end of the LAV envelope gene overlaps an open reading frame, termed F, that has the coding capacity for 206 amino acids and extends within the LTR (110 amino acids are encoded by the U3 region). The putatively encoded polypeptide (pF), the primary structure of which can be deduced, does not show any homology with the theoretical X gene products of the HTLV/BLV family. Also, the U5 and R elements are shorter (Table 2) and the polyadenylation signal is located within R, as is the case for all retroviruses except the HTLV/BLV. Additionally, LAV uses $tRNA^{lys}_3$ as (−) strand primer, as opposed to $tRNA^{pro}$ employed by all other mammalian retroviruses except MMTV (Donehower et al., 1981). Those homologies detected between the polymerase and protease domains of LAV and HTLV are also found in several retroviruses, RSV in particular.

It has been reported that a cloned HTLV-III genome hybridizes (T_m = 28°C) to sequences in the gag-pol and X regions of HTLV-I and -II; although restriction maps of cloned LAV and HTLV-III show almost perfect agreement (Hahn et al., 1984), we were unable to detect any such hybridization between LAV and HTLV-II (T_m = 55°C) (Alizon et al., 1984). Indeed, there is a punctual region of homology between LAV and HTLV-I (23/27' nucleotides starting at position 1859 in the LAV sequence) but nothing significant between the two viruses in the X region of HTLV-I. One possible reason for this discrepancy is that HTLV-III is subtly different from LAV. However it was subsequently reported that there was very minimal, if any, homology between orf X (of HTLV-I) and HTLV-III (Shaw et al., 1984).

Discussion

Regulatory sequences carried by retroviral LTR are believed to be involved in specific interactions between the viral genome and the host cell (Srinivasan et al., 1984). The LTR sequences of LAV are unique among retroviruses. That could reflect an original mode of gene expression, possibly in relation to particular transcriptional factors present in the virus-harboring cell. This hypothesis can be tested by studying the regulatory activity of the LAV LTR sequences in transient or long-term experiments involving an indicator gene and different cellular contexts.

The presence of the Q and F reading frames in addition to the conventional gag-pol-env set of genes is unexpected. One should now address the question of their role in the viral cycle and pathogenicity by trying to characterize their protein product(s). It is tempting to speculate on a role of such polypeptide(s) in T4 cells' mortality, a problem that can be studied by designing synthetic peptides for antibody production or by using site-directed mutagenesis of Q and F coding regions.

The peculiar genetic structure of LAV poses the question of its origin. The virus shares common tracts with other (apparently unrelated) retroviruses. For instance, the unusually large size of the outer membrane glycoprotein (env) and a comparably sized genome are also observed in the case of lentiviruses such as Visna (Harris et al., 1981; Querat et al., 1984). The presence of a large part of the F open reading frame in the LTR, and the use of $tRNA^{lys}_3$ as a primer for (−) strand synthesis, is reminiscent of the mouse mammary tumor virus. On the other hand, homologies in the pol gene would suggest that the LAV is closer to RSV than to any other retroviruses. Obviously, no clear picture can be drawn from the DNA sequence analysis as far as phylogeny is concerned. Thus, it may well be that LAV defines a new group of retroviruses that have been independently evolving for a considerable period of time, and not simply a variant recently derived from a characterized viral family. Both epidemiology and pathogeny of AIDS should be reconsidered with this idea in mind, when trying to answer such questions as these: Are there other human or animal diseases that are associated with similarly organized viruses? Is there a precursor to AIDS-associated virus(es) normally present, in latent form, in human populations? What triggered in this case the recent spreading of pathogenic derivatives?

Experimental Procedures

M13 Cloning and Sequencing

Total λJ19 DNA was sonicated, treated with the Klenow fragment of DNA polymerase plus deoxyribonucleotides (2 hr, 16°C), and fractionated by agarose gel electrophoresis. Fragments of 300–600 bp were excised, electroeluted, and purified by Elutip (Schleicher and Schüll) chromatography. DNA was ethanol-precipitated using 10 μg dextran T40 (Pharmacia) as carrier and ligated to dephosphorylated, Sma I-cleaved M13mp8 RF DNA using T4 DNA and RNA ligases (16 hr, 16°C) and transfected into E. coli strain TG-I. Recombinant clones were detected by plaque hybridization using the appropriate ^{32}P-labeled LAV restriction fragments as probes. Single-stranded templates were prepared from plaques exhibiting positive hybridization signals and were sequenced by the dideoxy chain termination procedure (Sanger et al., 1977) using α-^{35}S-dATP (Amersham, 400 Ci/mmol) and buffer gradient gels (Biggen et al., 1983). Sequences were compiled and analyzed using the programs of Staden adapted by B. Caudron for the Institut Pasteur Computer Center (Staden, 1982).

Strong-Stop cDNA

LAV virions from infected T lymphocyte (Barré-Sinoussi et al., 1983) culture supernatant were pelleted through a 20% sucrose cushion and the cDNA (−) strand was synthesized as described previously (Alizon et al., 1984) except that no exogenous primer was used. After alkaline hydrolysis (0.3 M NaOH, 30 min, 65°C), neutralization, and phenol extraction, the cDNA was ethanol-precipitated and loaded onto a 6%

acrylamide/8 M urea sequencing gel with sequence ladders as size markers.

Acknowledgments

We would like to thank Professors Luc Montagnier and Pierre Tiollais, in whose laboratory this work was carried out, for support and encouragement, as well as Professor Raymond Dedonder and Agnes Ullmann for their commitment to the project. Bernard Caudron and Jean-Noël Paulous of the Institut Pasteur Computer Center provided *invaluable and constant assistance, and Michelle Fonck, technical* support. Ana Cova and Louise-Marie Da tirelessly and good-humoredly typed the manuscript. We would like to thank Dr. Moshe Yaniv for critical reading of the manuscript and, finally, Genetic Systems, Seattle, WA, for communicating unpublished data.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received December 26, 1984

References

Alizon, M., Sonigo, P., Barré-Sinoussi, F., Chermann, J. C., Tiollais, P., Montagnier, L., and Wain-Hobson, S. (1984). Molecular cloning of lymphadenopathy-associated virus. Nature, in press.

Arya, S. K., Gallo, R. C., Hahn, B. H., Shaw, G. M., Popovic, M., Salahuddin, S. Z., and Wong-Staal, F. (1984). Homology of genome of AIDS-associated virus with genomes of human T-cell leukemia lymphoma viruses. Science *225*, 927–930.

Barré-Sinoussi, F., Chermann, J. C., Rey, F., Nugeybe, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vézinet-Brun F., Rouzioux, C., Rozenbaum, W., and Montagnier, L. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk of acquired immune deficiency syndrome (AIDS). Science *220*, 868–870.

Biggen, M. D., Gibson, T. J., and Hong, G. F. (1983). Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA *80*, 3963–3965.

Bird, A. P. (1980). DNA methylation and the frequency of CpG in animal DNA. Nucl. Acids Res. *8*, 1499–1504.

Brun-Vézinet, F., Rouzioux, C., Barré-Sinoussi, F., Klatzmann, D., Saimot, A. G., Rozenbaum, W., Montagnier, L., and Chermann, J. C. (1984). Detection of IgG antibodies to lymphadenopathy associated virus (LAV) by ELISA, in patients with acquired immuno-deficiency syndrome of lymphadenopathy syndrome. Lancet *I*, 1253–1256.

Chen, H. R., and Barker, W. C. (1984). Nucleotide sequences of the retroviral long terminal repeats and their adjacent regions. Nucl. Acids Res. *12*, 1767–1778.

Chen, I. S. Y., McLaughlin, J., Gasson, J. C., Clark, S. C., and Golde, D. W. (1983). Molecular characterization of the genome of a novel human T-cell leukaemia virus. Nature *305*, 502–505.

Chiu, I. M., Callahan, R., Tronick, S. R., Scholm, J., and Aaronson, S. A. (1984). Major pol gene progenitors in the evolution of oncornaviruses. Science *223*, 364–370.

Cianciolo, G. J., Kipnis, R. J., and Snyderman, R. (1984). Similarity between p15E of murine and feline viruses and p21 of HTLV. Nature *311*, 515.

Daly, H. M., and Scott, G. L. (1983). Fatal AIDS in a haemophiliac in the U.K. Lancet *II*, 1190.

Dittmar, K. J., and Moelling, K. (1978). Biochemical properties of p15-associated protease in an avian RNA tumor virus. J. Virol. *28*, 106–118.

Donehower, L. A., Huang, A. L., and Hager, G. L. (1981). Regulatory and coding potential of the mouse mammary tumour virus long terminal redundancy. J. Virol. *37*, 226–238.

Gottlieb, M. S., Schroff, R., Schanler, H. M., Weisman, J. D., Fan, P. T., Wolf, R. A., and Saxon, A. (1981). Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immuno-deficiency. N. Eng. J. Med. *305*, 1426–1431.

Hahn, B. H., Shaw, G. M., Arya, S. U., Popovic, M., Gallo, R. C., and Wong-Staal, F. (1984). Molecular cloning and characterization of the HTLV-III virus associated with AIDS. Nature *312*, 166–169.

Harris, J. D., Scott, J. V., Taylor, B., Brahic, M., Stowring, L., Ventura, P., Haase, A. T., and Peluso, R. (1981). Visna virus DNA: discovery of a novel gapped structure. Virology *113*, 573–583.

Kiyokawa, T., Yoshikura, H., Hattori, S., Secki, M., and Yoshida, M. (1984). Envelope proteins of human T-cell leukemia virus: expression in Escherichia coli and its application to studies of env gene functions. Proc. Natl. Acad. Sci. USA *81*, 6202–6206.

Klatzmann, D., Barré-Sinoussi, F., Nugeyre, M. T., Dauguet, C., Vilmer, E., Griscelli, C., Brun-Vézinet, F., Rouzioux, C., Gluckman, J. C., Chermann, J. C., and Montagnier, L. (1984). Selective tropism of lymphadenopathy associated virus (LAV) for helper-inducer T-lymphocytes. Science *225*, 59–63.

Kozak, M. (1984). Compilation and analysis of sequences upstream from the transcriptional start site in eucaryotic mRNAs. Nucl. Acids Res. *12*, 857–872.

Levy, J. A., Hoffman, A. D., Kramer, S. M., Lanois, J. A., Shimabukuro, J. M., and Oskiro, L. S. (1984). Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. Science *225*, 840–842.

Masur, H., Michelis, M. A., Greene, J. B., Onovato, I., Van de Stowe, R. A., Holzman, R. S., Wormser, G., Brettman, L., Lange, M., Murray, H. W., and Cunningham-Rundles, S. (1981). An outbreak of community-acquired pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction. N. Eng. J. Med. *305*, 1431–1438.

Misra, T. K., Grandgenett, D. P., and Parsons, J. T. (1982). Avian retrovirus pp32 DNA-binding protein. I. Recognition of specific sequences on retrovirus DNA terminal repeats. J. Virol. *44*, 330–343.

Montagnier, L., Chermann, J. C., Barré-Sinoussi, F., Chamaret, S., Gruest, J., Nugeyre, M. T., Rey, F., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Saimot, A. G., Rozenbaum, W., Gluckman, J. C., Klatzmann, D., Vilmer, E., Griscelli, C., Gazengel, C., and Brunet, J. B. (1984). A new human T-lymphotropic retrovirus: characterization and possible role in lymphadenopathy and acquired immune deficiency syndromes. In Human T-Cell Leukemia/Lymphoma Virus, R. C. Gallo, M. Essex, and L. Gross, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 363–370.

Oroszlan, S., Copeland, T. D., Kalyanaraman, V. S., Sarngadharan, M. G., Schultz, A. M., and Gallo, R. C. (1984). Chemical analysis of human T-cell leukemia virus structural proteins. In Human T-Cell Leukemia/Lymphoma Virus, R. C. Gallo, M. E. Essex, and L. Gross, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 101–110.

Piot, P., Quinn, T. C., Taelmann, H., Feinsod, F. M., Minlangu, K. B., Wobin, O., Mbendi, N., Mazebo, P., Ndangi, K., Stevens, W., Kalambayi, K., Mitchell, S., Bridts, C., and McCormick, J. B. (1984). Acquired immunodeficiency syndrome in a heterosexual population in Zaire. Lancet *II*, 65–69.

Popovic, M., Sarngadharan, M. G., Read, E., and Gallo, R. C. (1984). Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. Science *224*, 497–500.

Querat, G., Barban, N., Sauze, N., Filippi, P., Vigne, R., Russo, P., and Vitu, C. (1984). Highly lytic and persistent lentiviruses naturally present in sheep with progressive pneumonia are genetically distinct. J. Virol. *52*, 672–679.

Raba, M., Limburg, K., Burghagen, M., Katze, J. R., Simsek, M., Heckman, J. E., Rajbhandary, U. L., and Gross, H. J. (1979). Nucleotide sequence of three isoaccepting lysine tRNAs from rabbit liver and SV40-transformed mouse fibroblasts. Eur. J. Biochem. *97*, 305–318.

Rice, N. R., Stephen, R. M., Couez, D., Deschamps, J., Kettmann, R., Burny, A., and Gilden, R. V. (1984). The nucleotide sequence of the env gene and post-env region of bovine leukemia virus. Virology *138*, 82–93.

Sagata, N., Yasunaga, T., Ogawa, Y., Tsuzuku-Kawamura, J., and Ikawa, Y. (1984). Bovine leukemia virus: unique structural features of its long terminal repeats and its evolutionary relationship to human T-cell leukemia virus. Proc. Natl. Acad. Sci. USA *81*, 4741–4745.

Sanger, F., Nicklen, S., and Coulsen, A. R. (1977). DNA sequencing

with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463–5467.

Schüpbach, J., Popovic, M., Gilden, R. V., Gonda, M. A., Sarngadharan, M. G., and Gallo, R. C. (1984). Serological analysis of a subgroup of human T-lymphotropic retroviruses (HTLV-III) associated with AIDS. Science *224*, 503–505.

Schwartz, D. E., Tizard, R., and Gilbert, W. (1983). Nucleotide sequence of Rous sarcoma virus. Cell *32*, 853–869.

Seiki, M., Hattori, S., Hirayama, Y., and Yoshida, M. (1983). Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. Proc. Natl. Acad. Sci. USA *80*, 3618–3622.

Shaw, G. M., Hahn, B. H., Arya, S. K., Groopman, J. E., Gallo, R. C., and Wong-Staal, F. (1984). Molecular characterization of human T-cell leukemia (lymphotropic) virus type III in the acquired immune deficiency syndrome. Science *226*, 1165–1171.

Shimotohno, K., and Temin, H. M. (1982). Spontaneous variation and synthesis in the U3 region of the long terminal repeat of an avian retrovirus. J. Virol. *41*, 163–171.

Shimotohno, K., Golde, D. M., Miwa, M., Sugimura, T., and Chen, I. S. Y. (1984). Nucleotide sequence analysis of the long terminal repeat of human T-cell leukemia virus type II. Proc. Natl. Acad. Sci. USA *81*, 1079–1083.

Shinnick, T. M., Lerner, R. A., and Sutcliffe, J. G. (1981). Nucleotide sequence of Moloney murine leukemia virus. Nature *293*, 543–548.

Srinivasan, A., Reddy, E. P., Dunn, C. Y., and Aaronson, S. A. (1984). Molecular dissection of transcriptional control elements with the long terminal repeat of retrovirus. Science *223*, 286–289.

Staden, R. (1982). Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. Nucl. Acids. Res. *10*, 4731–4751.

Temin, H. (1981). Structure, variation and synthesis of retrovirus long terminal repeat. Cell *27*, 1–3.

Weinberg, R. A. (1982). Fewer and fewer oncogenes. Cell *30*, 3–9.

Proceedings
OF THE
National Academy of Sciences
OF THE UNITED STATES OF AMERICA

June 1983 | Volume 80, Number 12 | pp. 3543–3878

Table of Contents

Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 3618–3622, June 1983
Biochemistry

Human adult T-cell leukemia virus: Complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA

(human leukemia virus/provirus structure/translation frames/polyadenylylation model)

MOTOHARU SEIKI, SEISUKE HATTORI, YOKO HIRAYAMA, AND MITSUAKI YOSHIDA

Department of Viral Oncology, Cancer Institute, Kami-Ikebukuro, Toshima-ku, Tokyo, Japan

Communicated by Takashi Sugimura, March 14, 1983

ABSTRACT Human retrovirus adult T-cell leukemia virus (ATLV) has been shown to be closely associated with human adult T-cell leukemia (ATL) [Yoshida, M., Miyoshi, I. & Hinuma, Y. (1982) *Proc. Natl. Acad. Sci. USA* 79, 2031–2035]. The provirus of ATLV integrated in DNA of leukemia T cells from a patient with ATL was molecularly cloned and the complete nucleotide sequence of 9,032 bases of the proviral genome was determined. The provirus DNA contains two long terminal repeats (LTRs) consisting of 755 bases, one at each end, which are flanked by a 6-base direct repeat of the cellular DNA sequence. The nucleotides in the LTR could be arranged into a unique secondary structure, which could explain transcriptional termination within the 3′ LTR but not in the 5′ LTR. The nucleotide sequence of the provirus contains three large open reading frames, which are capable of coding for proteins of 48,000, 99,000, and 54,000 daltons. The three open frames are in this order from the 5′ end of the viral genome and the predicted 48,000-dalton polypeptide is a precursor of gag proteins, because it has an identical amino acid sequence to that of the NH_2 terminus of human T-cell leukemia virus (HTLV) p24. The open frames coding for 99,000- and 54,000-dalton polypeptides are thought to be the *pol* and *env* genes, respectively. On the 3′ side of these three open frames, the ATLV sequence has four smaller open frames in various phases; these frames may code for 10,000-, 11,000-, 12,000-, and 27,000-dalton polypeptides. Although one or some of these open frames could be the transforming gene of this virus, in preliminary analysis, DNA of this region has no homology with the normal human genome.

Recently, retroviruses were independently isolated from human T-cell leukemias by two groups. One retrovirus is human T-cell leukemia virus (HTLV) isolated by Gallo and colleagues from patients with cutaneous T-cell lymphoma (1, 2), and the other is adult T-cell leukemia virus (ATLV) isolated from patients with adult T-cell leukemia (ATL) (3, 4). Recently, these two viruses have been shown to be closely related (5). ATLV was shown to be associated with ATL, which is a unique disease with T-cell malignancy (6), and the provirus genome was always detected in the chromosomal DNA of the leukemia cells (4). Recently, we reported molecular cloning of provirus DNA integrated in the cell line MT-1 and the nucleotide sequence of the long terminal repeat (LTR) with 754 bases (7), and we also proposed that ATLV might be distinct from other known animal retroviruses (7). From these previous observations, identification of genetic structure and the gene products seemed to be of great importance in understanding the origin of the virus and the mechanisms of leukemogenesis by this virus. For this purpose, we isolated a clone (λATK-1) of the provirus genome integrated in ATL cell DNA.

This paper reports the complete 9,032-nucleotide sequence of the proviral genome cloned in λATK-1 and the amino acid sequence predicted for the putative proteins.

MATERIALS AND METHODS

Cloning and Sequence Analysis of Provirus DNA of ATLV Integrated in Leukemia Cells. DNA was extracted from peripheral blood cells of a patient (K. K.) with ATL, digested with *Eco*RI, and separated by electrophoresis in agarose gel. DNA fractions of the 17-kilobase fragment containing the provirus were extracted, ligated to the *Eco*RI site of Charon 4A phage DNA, and subjected to *in vitro* packaging as described by Blattner *et al.* (8). Screening with viral [^{32}P]cDNA, recombinant phage λATK-1 was isolated. The DNA fragment cloned in λATK-1 was excised by *Eco*RI and cleaved into several fragments with restriction endonucleases for subcloning in plasmid pBR322. The nucleotide sequence of the fragments was determined by the procedure of Maxam and Gilbert (9).

RESULTS

Molecular Cloning and Sequence Analysis Strategy. Previously we reported the molecular cloning (λATM-1) of the provirus genome from cell line MT-1 and identified the LTR structure (7). However, this time we have isolated a new provirus clone λATK-1 directly from DNA of leukemia cells of an ATL patient for further analysis.

A simple restriction cleavage map of the inserted fragment in λATK-1 was constructed to subclone the regions containing provirus into pBR322. As shown in Fig. 1, *Bam*HI divided the viral sequence into three fragments and these were subcloned into pBR322; thus, pATK-03, pATK-06, and pATK-08 were obtained. Plasmid pATK-100, constructed from the *Pst* I fragment of the λATK-1 insert, contained two *Bam*HI junctions between the subclones described above. The plasmids pATK-03, pATK-06, and pATK-08 were digested with *Pst* I, *Sal* I, and *Sma* I, respectively, and the fragments were subjected to sequence analysis in both strands after further digestions with *Hpa* II, *Sau*3AI, *Hin*fI, or other restriction endonucleases. The determined sequences of pATK-03, pATK-06, and pATK-08 were overlapped by sequence analysis across the two *Bam*HI sites in the clone pATK-100. Fig. 2 shows the 9,032-nucleotide sequence of the constructed whole provirus genome with two LTRs, together with the cellular flanking sequences.

DISCUSSION

Provirus Structure. The LTR structure (U3-R-U5) is thought to play essential roles in integration of provirus DNA into the host chromosomal DNA and also in regulation of transcription of the provirus genome (10, 11). The provirus DNA in λATK-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: ATL, adult T-cell leukemia; ATLV, ATL virus; HTLV, human T-cell leukemia virus; LTR, long terminal repeat.

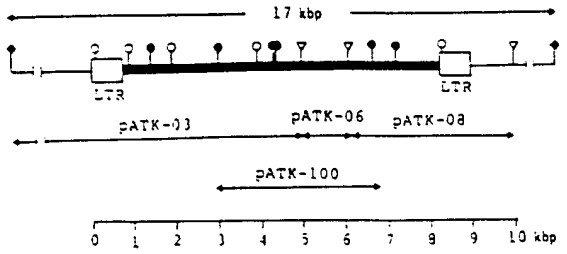


FIG. 1. Restriction map of ATLV provirus clones. The provirus DNA is shown by the thick line with a LTR (box) at each end. The positions of the inserts from clones pATK-03, pATK-06, pATK-08, and pATK-100 are shown under the full provirus genome in λATK-1. ◆, *Eco*RI; ▷, *Sma* I; ●, *Pst* I; and ▽, *Bam*HI. kbp, Kilobase pairs.

1 contained two direct repeats of the LTR sequence, one at each end, and the structural features were similar to those in λATM-1, which was isolated from cell line MT-1 (7). Comparison of these two clones revealed the following features. (*i*) Sequences of the LTRs are identical except for 6 base changes at positions 38, C to T; 90, G to A; 146, A to G; 209, G to A; 316, A to G; 481, G to A; and one base (A) insertion at position 190. (*ii*) Cellular flanking sequences are directly repeated by 6 bases in both clones, but the sequences themselves are different, reflecting different integration sites (Fig. 3). Previously, we reported 7-base direct repeats of cellular sequences in λATM-1, but careful reinvestigation demonstrated that there are in fact 6-base repeats. (*iii*) The lengths of the viral sequences between the two LTRs are identical within the limits of experimental errors, although the nucleotide sequence of λATM-1 was not fully determined. The above results indicate that two clones, from cell line and leukemia blood cells, represent a similar ATLV genome.

The unique structures of the LTR previously reported (7) have also been confirmed in this paper. These are (*i*) the extremely long size of R (terminally redundant sequence of genomic RNA) with 229 bases and (*ii*) the absence of the poly(A) signal around the poly(A) site, which is the end of R. With few exceptions, all eukaryotic mRNA containing poly(A) contained the poly(A) signal A-A-T-A-A-A at 10–30 bases upstream of the poly(A) site, but from the sequence of ATLV LTR, we speculated in the previous paper (7) that the poly(A) signal is dispensable for polyadenylylation. However, the nucleotide sequence in the LTR was found to be arranged into a possible secondary structure (Fig. 4), which may explain why transcription terminates within the 3′ LTR but does not terminate in the 5′ LTR. In the 3′ LTR, the RNA transcript that had been initiated at the 5′ LTR would form a hairpin structure, as shown in Fig. 4; thus, the poly(A) signal A-A-T-A-A-A, which is located before the "TATA" box or at 276 bases upstream of the poly(A) site, is arranged into 20 bases before the poly(A) site. In this structure, the signal A-A-T-A-A-A might become effective in the RNA level. But in the 5′ LTR, transcription starts from the cap site, which is located in the loop; therefore, the RNA transcript lacks the poly(A) signal, thus allowing further transcription. A model for inactivation of the A-A-T-A-A-A signal by a possible secondary structure was also proposed in the LTR of murine leukemia virus by Benz *et al.* (12). Our model for ATLV suggests that signals separated by a long nucleotide sequence could be aligned into functional form by conformational rearrangements; therefore, a definite structure in the primary sequences might not necessarily be required. However, this could be an exceptional case.

Capacity of the Genome To Code the Proteins. In general, replication-competent retroviruses have a common gene organization that is *gag*, *pol*, and *env* in this order from the 5′ end of the genomic RNA (13). The DNA sequence of ATLV contained three large open reading frames and four additional smaller ones (Fig. 2). Other possible open frames in the various phases are <200 bases, corresponding to a coding capacity for 70 amino acids. The three large reading frames probably correspond to *gag*, *pol*, and *env* because of their positions and for reasons discussed later.

gag *gene*. The first open frame, which starts from the ATG codon at position 802 and terminates with TAA at position 2,089, could code for a 48,000-dalton protein consisting of 429 amino acids. The recently reported NH_2-terminal sequence of 25 amino acids of p24 in HTLV (14), which is similar to ATLV (5), is identical to a part of this 48,000-dalton protein, which starts from proline at position 1,192, as marked in Fig. 2. The COOH terminus of p24 of HTLV is leucine (14) and this may correspond to the leucine at position 1,831. The predicted p24 of ATLV has a molecular mass of 23,940 daltons and its amino acid composition is very similar to that of p24 of HTLV reported by Oroszlan *et al.* (Table 1) (14). This finding is direct evidence that p24 is virus encoded and also is consistent with the fact that an antibody against p24 of HTLV is crossreactive with ATLV antigens (15). Thus, the first large open frame appears to be the *gag* gene coding for a *gag*-precursor protein, $Pr48^{gag}$. To form p24, the $Pr48^{gag}$ should be cleaved into at least three proteins—that is, a 14,000-dalton protein from the NH_2-terminal, a 24,000-dalton protein from the middle, and a 9,000-dalton protein from the COOH terminal portions of the $Pr48^{gag}$. The molecular masses of the presumed polypeptides may correspond to the 17,000-, 24,000-, and 11,000-dalton proteins, within the limits of experimental errors; these proteins were found previously to be associated with ATLV virions (4).

pol *gene*. In animal retroviruses, the *pol* gene is located after the *gag* gene and is translated into the *gag–pol* polyprotein by changing the reading frame after splicing of the genomic RNA (ref. 16) or by suppressing one termination codon, which appears after the *gag* gene in the frame (17). Because ATLV has the general structural features of the retrovirus genome, such as LTR structure and tRNA binding site (7), it is reasonable to expect that ATLV has the usual gene organization. Thus, the second reading frame from GCC at position 2,498 to TAA at position 5,185 is expected to be the *pol* gene coding for reverse transcriptase. This is the largest open frame and it can code for

Table 1. Amino acid composition of p24

Amino acid	p24 of ATLV	p24 of HTLV*
Asn	9	(21†
Asp	10	
Thr	9	10
Ser	13	14
Gln	21	(36‡
Glu	9	
Pro	18	22
Gly	11	15
Ala	20	24
Cys	3	—
Val	9	7
Met	4	4
Ile	8	8
Leu	28	32
Tyr	5	6
Phe	4	5
His	8	9
Lys	10	12
Arg	11	11
Trp	4	—

* Oroszlan *et al.* (14).
† Asn and Asp.
‡ Gln and Glu.

5'-LTR
gag
pol

FIG. 2. Complete nucleotide sequence of the ATLV provirus genome. The nucleotides are numbered from the 5′ end of the 5′ LTR. The sequence is translated into amino acids in the region from positions 802 to 2,088 and positions 2,498 to 6,043, and the remaining smaller open frames are marked by small vertical lines. All open frames are started by ATG, except the putative *pol* gene, which started with GGC. [ ], LTR; □, TATA box; ——, poly(A) signal; ------, the sequence R; and ■, the NH_2- and COOH-terminal amino acids reported for p24.

FIG. 3. Nucleotide sequences of the virus–cellular junction in the two clones λATK-1 and λATM-1.

896 amino acids, corresponding to a 99,000-dalton protein. This molecular mass is similar to that of the known reverse transcriptase, but we could not define the NH_2 terminus, because no structural information on the enzyme of ATLV or HTLV is available. Because there are several termination codons in every reading frame after the *gag* gene [at positions 2,089, 2,161, 2,182, 2,239, 2,257, 2,272, 2,347, 2,422, 2,455, and 2,495 in the frame for *gag* and *pol* (frame I), positions 2,123, 2,186, 2,198, 2,258, and 2,438 in frame II, and positions 2,316, 2,370, 2,466, 2,418, and 2,448 in frame III], splicing of the genomic RNA is expected to eliminate the stop codons to read through *gag* to the putative *pol* gene, although we have no evidence for a possible presence of a polyprotein of *gag–pol*.

env gene. The third large open frame, which starts at the ATG codon at position 5,180 and terminates with the TAA codon at position 6,644, has the capacity to code for a 54,000-dalton protein composed of 488 amino acids. This frame and the predicted amino acids have the following features in common with the *env* gene products of animal retroviruses. (*i*) The ATG codon at position 5,180 for initiation of the 54,000-dalton protein is located within the putative *pol* gene overlapping by 5 bases. Similar overlappings between *pol* and *env* are also observed in Rous sarcoma virus (D. Schwarz, R. Tizard, and W. Gilbert, personal communication) and murine leukemia virus genomes (18). (*ii*) About 20 amino acids of the NH_2-terminal portion are rich in hydrophobic residues, and this characteristic is similar to that of signal peptides proposed for the *env* gene product of Rous sarcoma virus and murine leukemia virus (18). (*iii*) The 54,000-dalton protein contains five possible sites for glycosylation—that is, Asn-X-Thr/Ser sequences (19) at positions 5,597, 5,843, 5,909, 5,993, and 6,389. Because the *env* gene products are generally glycoproteins, presence of the sites for glycosylation is expected to be essential, although it may not be enough. The product of the *env* of ATLV or HTLV has not been identified, but the characteristics of the putative 54,000-dalton protein described above suggest that this open frame is the *env* gene rather than the *onc* gene.

Other genes? In addition to *gag*, *pol*, and *env*, the ATLV sequence determined has four extra open frames, as indicated in Fig. 2, which have capacities to code for proteins pX-I to pX-IV, with molecular masses of 11,000, 10,000, 12,000, and 27,000 daltons, respectively. Although the presence of these proteins in infected or leukemia cells remains to be studied, some of them might have functions in the process of transformation of infected T cells. If some of these sequences have the common features with the known *onc* genes in acute leukemia viruses, similar nucleotide sequences are expected to be present in normal human DNA. However, the subcloned DNA fragment containing this region did not significantly hybridize with normal human DNA in Southern blotting analysis. This preliminary result indicated that the region containing four extra open frames is not homologous with the human c-*onc* genes. Similar experiments using the other parts of viral DNA fragments suggested that ATLV has no *onc* gene derived from the human genome; however, it is possible that ATLV may contain a gene that is involved in induction of abnormal T-cell proliferation but not derived from the human DNA.

Finally, it should be pointed out that the predicted viral genes or gene products could be tentative, because the provirus analyzed in this paper is that integrated in leukemia cells, and we have no direct evidence for the replicative competence of this provirus, including the viral infection.

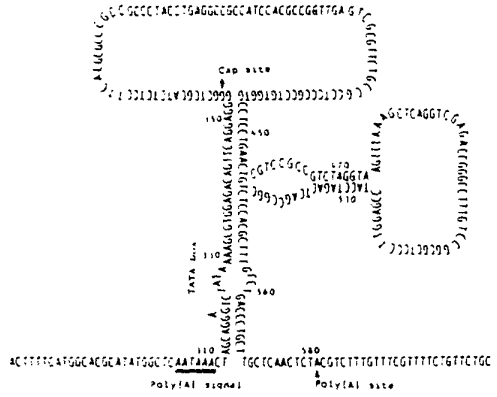


FIG. 4. Possible secondary structure of the nucleotide sequence around the cap site and poly(A) site in the LTR.

The authors thank Dr. H. Sugano for valuable discussion and encouragement during this work. This work was supported in part by a Grant-in-Aid for Cancer Research from the Ministry of Education, Science and Culture of Japan.

1. Poiesz, B. J., Ruscetti, F. W., Gazdar, A. F., Bunn, P. A., Minna, J. D. & Gallo, R. C. (1980) *Proc. Natl. Acad. Sci. USA* 77, 7415–7419.
2. Reitz, M. S., Poiesz, B. J., Ruscetti, F. W. & Gallo, R. C. (1981) *Proc. Natl. Acad. Sci. USA* 78, 1887–1891.
3. Hinuma, Y., Nagata, K., Hanaoka, M., Nakai, M., Matsumoto, T., Kinoshita, K., Shirakawa, S. & Miyoshi, I. (1981) *Proc. Natl. Acad. Sci. USA* 78, 6476–6480.
4. Yoshida, M., Miyoshi, I. & Hinuma, Y. (1982) *Proc. Natl. Acad. Sci. USA* 79, 2031–2035.
5. Popovic, M., Reitz, M. S., Sarngadharan, M. G., Robert-Guroff, M., Kalyanaraman, V. S., Nakao, Y., Miyoshi, I., Minowada, J., Yoshida, M., Ito, Y. & Gallo, R. C. (1982) *Nature (London)* **300**, 63–66.
6. Uchiyama, T., Yodoi, J., Segawa, K., Takatsuki, K. & Uchino, H. (1977) *Blood* **50**, 481–492.
7. Seiki, M., Hattori, S. & Yoshida, M. (1982) *Proc. Natl. Acad. Sci. USA* 79, 6899–6902.
8. Blattner, F. R., Blechl, A., Denniston-Thompson, K., Faber, H. E., Richards, J. E., Slightom, J. L., Tucker, P. W. & Smithies, O. (1978) *Science* **202**, 1279–1284.
9. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
10. Hsu, T. W., Sabaran, J. L., Mark, G. E., Guntaka, R. V. & Taylor, J. M. (1978) *J. Virol.* **28**, 810–818.
11. Varmus, H. E. (1982) *Science* **216**, 812–820.
12. Benz, E. W., Jr., Wydro, R. M., Nadal-Ginard, B. & Dina, D. (1981) *Nature (London)* **288**, 665–669.
13. Vogt, P. K. (1977) *Compr. Virol.* **10**, 341–455.
14. Oroszlan, S., Sarngadharan, M. G., Copeland, T. D., Kalyanaraman, V. S., Gilden, R. V. & Gallo, R. C. (1982) *Proc. Natl. Acad. Sci. USA* 79, 1291–1294.
15. Kalyanaraman, V. S., Sarngadharan, M. G., Nakao, Y., Ito, Y., Aoki, T. & Gallo, R. C. (1982) *Proc. Natl. Acad. Sci. USA* 79, 1653–1657.
16. Weiss, S. R., Hackett, P. B., Opperman, H., Ullrich, A., Levinthow, L. & Bishop, J. M. (1978) *Cell* **15**, 607–614.
17. Philipson, L., Andersson, P., Olshevsky, U., Weinberg, R. & Baltimore, D. (1978) *Cell* **13**, 189–199.
18. Shinnick, T. M., Lerner, R. A. & Sutcliffe, J. G. (1981) *Nature (London)* **293**, 543–548.
19. Chen, R. (1982) *Proc. Natl. Acad. Sci. USA* 79, 5788–5792.

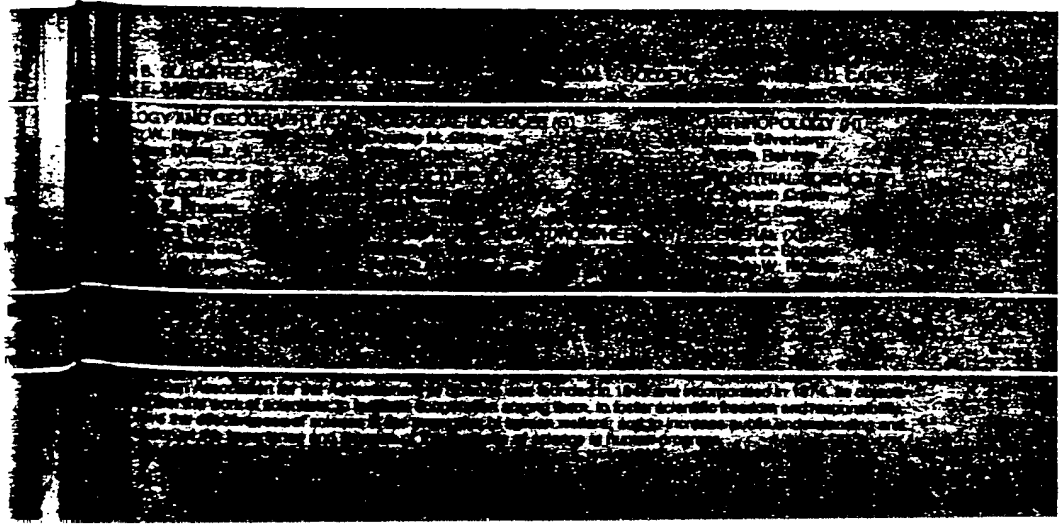
ISSN 0036-8075

27 July 1984

Volume 225, No. 4660

12

...ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE



COVER

Color-classified Seasat synthetic aperture radar image of pack ice in the Beaufort Sea west of Banks Island, Northwest Territories, Canada (4 October 1978). The image is a combination of the color-classified image and the original image and shows the following separable ice classes: red, multiyear ice; black, new or grease ice; yellow, young or pancake ice; and bluish-white, open water. See page 371. [W. F. Weeks, Snow and Ice Branch, Cold Regions Research and Engineering Laboratory, Hanover, New Hampshire 03755]

host factors encoded by dominant alleles at the *Fv*-1 locus (*13*).

To our knowledge, these results are the first report of a viral capsid protein playing a critical role in the congenital transmission of a retrovirus. Whether capsid proteins affect the replication of other families of retroviruses in reproductive tissue is not known. However, since the ability to undergo efficient congenital transmission has survival value for exogenous but not endogenous viruses, the major capsid proteins for all exogenous and endogenous viruses may have undergone selection for their ability to ensure or restrict the replication of virus in reproductive tissue. If so, the capsid proteins of exogenous and endogenous viruses may provide genes that can be used to construct viruses that either will or will not undergo congenital transmission.

HARRIET L. ROBINSON
Worcester Foundation for Experimental Biology, Shrewsbury, Massachusetts 01545
ROBERT N. EISENMAN
Fred Hutchinson Cancer Research Center, Seattle, Washington 98104

References and Notes

1. H. Rubin, A. Cornelius, L. Fanshier, *Proc. Natl. Acad. Sci. U.S.A.* 47, 1058 (1962); B. R. Burmester, R. F. Gentry, N. F. Waters, *Poult. Sci.* 34, 609 (1955).
2. R. M. Dougherty and H. S. DiStefano, *Cancer Res.* 27, 322 (1967).
3. H. Rubin, L. Fanshier, A. Cornelius, W. F. Hughes, *Virology* 17, 143 (1962).
4. H. L. Robinson and W. F. Lamoureux, *ibid.* 69, 50 (1976); H. L. Robinson, S. M. Astrin, A. M. Senior, F. H. Salazar, *J. Virol.* 40, 745 (1981).
5. H. L. Robinson, B. M. Blais, P. N. Tsichlis, J. M. Coffin, *Proc. Natl. Acad. Sci. U.S.A.* 79, 1225 (1982).
6. H. L. Robinson, M. N. Pearson, D. W. DeSimone, P. N. Tsichlis, J. M. Coffin, *Cold Spring Harbor Symp. Quant. Biol.* 44, 1133 (1979).
7. P. N. Tsichlis and J. M. Coffin, *J. Virol.* 33, 238 (1980).
8. J. M. Coffin, P. N. Tsichlis, K. F. Conklin, A. Senior, H. L. Robinson, *Virology* 126, 51 (1983).
9. R. J. Huebner, D. Armstrong, M. Okuyan, P. S. Sarma, H. C. Turner, *Proc. Natl. Acad. Sci. U.S.A.* 51, 742 (1964).
10. R. A. Weiss, *J. Gen. Virol.* 5, 511 (1969); *ibid.*, p. 529.
11. C. W. Rettenmier and H. Hanafusa, *J. Virol.* 24, 850 (1977); R. Shaikh, M. Linial, J. Coffin, R. Eisenman, *Virology* 87, 326 (1978).
12. J. Schindler, R. Hynes, N. Hopkins, *J. Virol.* 23, 700 (1977); D. V. Faller and N. Hopkins, *ibid.* 26, 153 (1978); J. W. Gautsch, J. H. Elder, J. Schindler, F. C. Jensen, R. A. Lerner, *Proc. Natl. Acad. Sci. U.S.A.* 75, 4170 (1978).
13. T. Pincus, J. W. Hartley, W. P. Rowe, *J. Exp. Med.* 133, 1219 (1971); F. Lilly and T. Pincus, *Adv. Cancer Res.* 17, 231 (1973).
14. R. N. Eisenman and V. M. Vogt, *Biochim. Biophys. Acta* 473, 187 (1978).
15. H. L. Robinson, unpublished observation.
16. We thank B. Blais, G. Gagnon, and S. Johnson for technical assistance; H. Hanafusa and T. Hanafusa for the RAV-60 viruses; J. Coffin and P. Tsichlis for NTRE-7 and NTRE-2; P. Tsichlis for pRAV-0; G. Payne, M. Bishop, and H. Varmus for pRAV-1; and D. Steffen and S. Wadsworth for comments on the manuscript. Supported by NIH grants R01 CA23086 and P30 12708 and by the W. J. Tannenberg fund (H.L.R.) and NIH grant R01 CA20525 (R.N.E.). R.N.E. is a scholar of the Leukemia Society of America.

12 March 1984; accepted 4 May 1984

Structure of 3′ Terminal Region of Type II Human T Lymphotropic Virus: Evidence for New Coding Region

Abstract. *The sequence of the 3′ terminus of the human T lymphotropic virus type II (HTLV-II) was determined and compared to the corresponding sequence of HTLV-I. The 1557-nucleotide-long sequence can be divided into a 5′ region that is not conserved between the two viruses, and a 3′, 1011-nucleotide-long region that is highly conserved and that corresponds precisely with a long open reading frame for both HTLV-I and -II. The proteins that could be encoded by these open reading frames have a molecular weight of about 38,000 and are closely related in primary amino acid sequence. The genomic structure in the 3′ region of HTLV was found to be similar to that of bovine leukemia virus.*

The human T lymphotropic viruses (HTLV) are a family of retroviruses that are associated with T-cell abnormalities (*1*). Isolates known as HTLV-I are associated with an aggressive form of adult T-cell leukemia or lymphoma (*1*). An infrequent isolate known as HTLV-II was first identified in a patient with a T-cell variant of hairy cell leukemia (*2*). Recently, some viruses collectively called HTLV-III were isolated from patients with the acquired immune deficiency syndrome (*3*).

The genomes of HTLV-I and -II differ from those of the nonacute retroviruses, which encode only the *gag*, *pol*, and *env* genes, in that they have an additional sequence that is approximately 1600 nucleotides long. This sequence is located between the 3′ end of the *env* gene and the 5′ end of the U3 region of the proviral long terminal repeat (LTR) (*4*).

Although this sequence occupies a position similar to the *src* gene in Rous sarcoma virus, it is not homologous to conserved mammalian genes and therefore differs from the oncogenes of transforming retroviruses (*4*). There is some evidence that this region contains a functional gene. Heteroduplex analysis of HTLV-I and -II reveals a conserved sequence about 1000 nucleotides long near the 3′ terminus of the genome (*5*). Spliced messenger RNA (mRNA) species that contain sequences that are unique to the 5′ end of the viral genome (U5 LTR sequences) and a portion of the 3′ sequence are observed in HTLV-infected cells and in some fresh tumor cells (*6*). Seiki *et al.* (*4*) note that several open reading frames occur within the 3′ sequence of HTLV-I.

To obtain a clearer understanding of the potential role of the 3′ region of HTLV, we determined the primary nucleotide sequence of the region located between the 3′ end of the *env* gene and the LTR of a cloned HTLV-II provirus, MO15A (*7*).

The nucleotide sequence of 1557 bases of the 3′ terminal region of HTLV-II is presented in Fig. 1. This sequence can be divided into two regions. One region, 546 nucleotides long, is located at the 5′ end of the sequence and has either no or very little similarity to the corresponding sequences in HTLV-I. For this reason we call this sequence the nonconserved region (NCR). A second region, 1011 nucleotides long, comprises the 3′ portion of this sequence. This sequence is very similar to that of HTLV-I and is identical at 765 of 1011 nucleotides (76 percent identity).

A new gene? The perimeters of the 1011 nucleotide sequence of the HTLV-II genome correspond precisely with a single long open reading frame capable of encoding a polypeptide 337 amino acids long. A corresponding sequence of HTLV-I also encompasses a single long open reading frame capable of encoding a polypeptide 357 amino acids long. We call the nucleotide sequence containing these long open reading frames the LOR region (nucleotides 566 to 1557 in HTLV-II) (Fig. 1).

The predicted amino acid sequences of both polypeptides are presented in Fig. 1. The potential proteins encoded by the LOR regions of HTLV-I and -II are of approximately the same length and are identical in 259 of 337 of the amino acids (77 percent identity). The degree of similarity of these two proteins is even more striking if conservative amino acid substitutions are considered (89 percent similar). The distribution of hydrophilic and hydrophobic regions of these proteins is remarkably similar (Fig. 2).

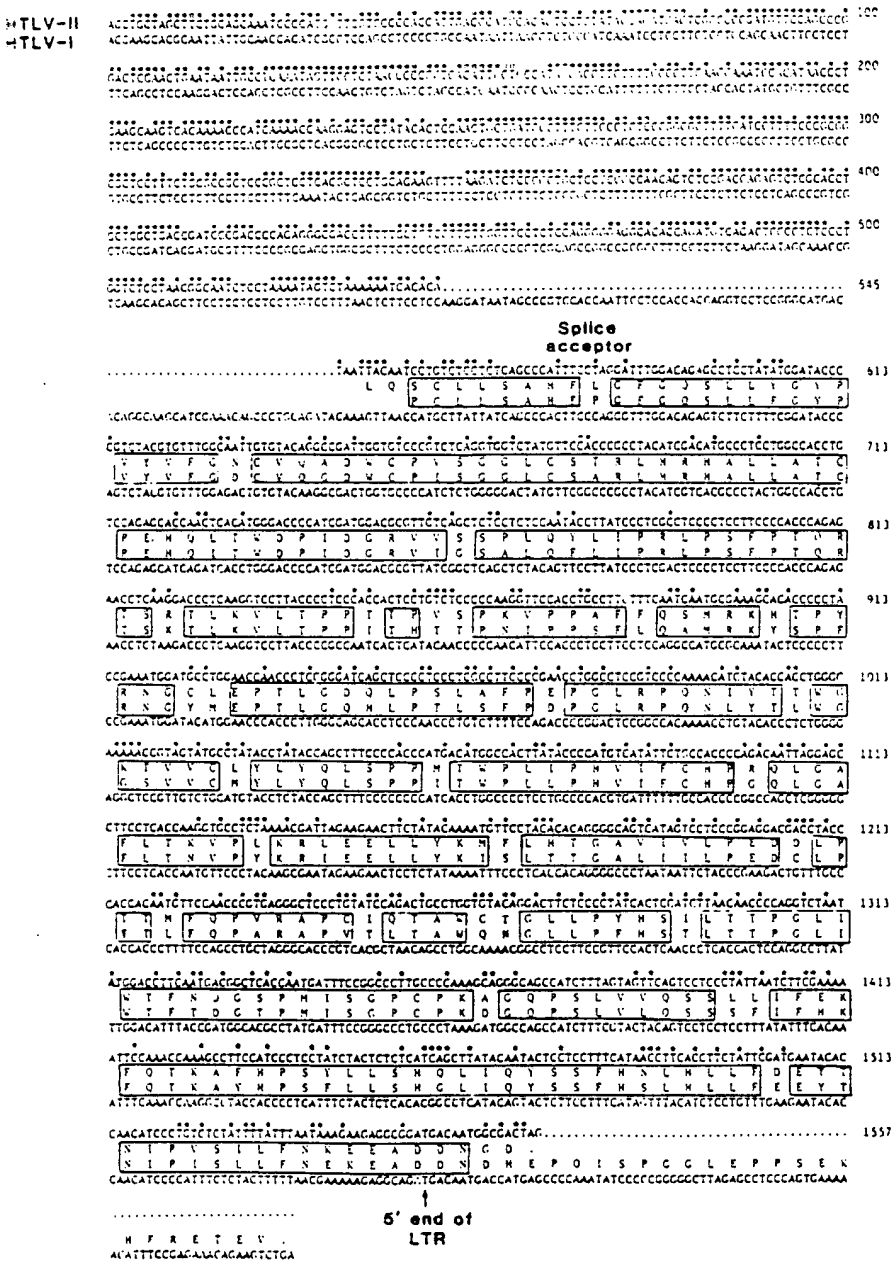
We also note the existence of a splice acceptor consensus sequence located at the 5′ end of the open reading frame (Fig. 1). Although no methionine codon occurs at the 5′ end of the open reading frames of HTLV-I and -II, a fusion protein synthesized from a spliced mRNA can be envisioned. Several other splice acceptor sequences occur within this reading frame from which smaller fusion proteins might also be made.

These observations suggest that the 3′ terminal region of HTLV contains a new gene that encodes a protein with a mo-

lecular weight of at least 38,000. Such a protein could be translated from the 2.2-kb spliced mRNA species containing LOR sequences found in HTLV-infected cells (*6*). A protein of molecular weight 38,000 to 42,000 in HTLV-I–infected cell lines has been noted that is recognized by the serum of persons infected with HTLV-I, but not by serum from control subjects (*8*).

Several other open reading frames exist in the region between the *env* gene and the LTR of both HTLV-I and -II. Seiki *et al.* (*4*) have identified four such regions, pX I to pX IV. The pX IV region corresponds to the carboxyl terminus of the peptide that could be encoded by the LOR region. No region of predicted protein similarity could be found in HTLV-II that corresponds to pX I or pX III. A further argument against the functional importance of pX I is that an 11-nucleotide deletion that destroys the pX I open reading frame occurs in an HTLV-Ic isolate with apparently complete biological activity (*9*). Another open reading frame in the LOR region of HTLV-II (nucleotides 530 to 1325) includes a region exhibiting 65 percent amino acid homology to pX II. The significance of this similarity is not clear, because the pX II peptide is much shorter than the corresponding peptide in HTLV-II (87 compared to 265 amino acids). Sequence similarity here could arise as a result of conservation of the LOR protein in the other open reading frame.

We have also reported (*8*) that *trans*-acting factors, either directly encoded by



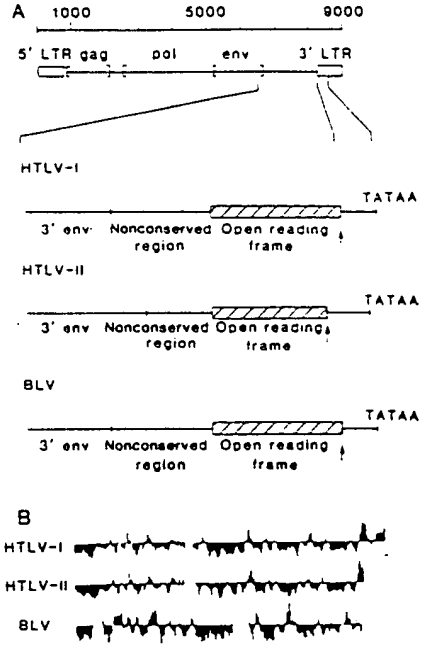


Fig. 1 (left). Nucleotide sequence of the HTLV-II 3' terminal region and predicted amino acid sequence of its potential product. Plasmid DNA containing the 3' portion of MO15A, an HTLV-II proviral clone (*7*), was cleaved with either Cla I or Bgl II, which cut the plasmid uniquely at a single site. After timed digestion with Bal 31 exonuclease, the ends were blunted with T4 DNA polymerase and synthetic linkers were added prior to recloning. Linker sites separated by increments of 100 to 200 nucleotides were end-labeled and the fragments sequenced by the method of Maxam and Gilbert (*19*). The sequence of the 3' region following the termination codon for the envelope gene is presented for HTLV-II and HTLV-I. The HTLV-II sequence is numbered according to the nucleotides following the envelope stop codon. Asterisks represent differences between the DNA sequences. The positions of a conserved splice acceptor consensus sequence and the 5' end of the LTR are noted. Note that the sequence is not well conserved 5' to the putative splice acceptor site but is very well conserved 3' to this site. The latter sequence corresponds to a long open reading frame (LOR) region. The predicted amino acid sequences of the potential products of the HTLV-I and HTLV-II 3' open reading frames are optimally aligned. Boxed regions indicate amino acid identity or conservative amino acid substitutions between the sequences. Fig. 2 (right). The open reading frames of the HTLV and BLV genomes. (A) The position of 3' open reading frames in the genomes of HTLV-I and -II and of BLV. The 3' end of the envelope gene is shown, as well as the 5' terminus of the LTR (↑) and the promoter (TATAA) sequence. The positions of the nonconserved regions and the open reading frames (hatched boxes) are displayed. (B) The relative hydrophilicity of the 3' open reading frame products of HTLV-I, HTLV-II, and BLV calculated according to the method of Hopp and Woods (*20*). Hydrophilic regions are shown above the axis, hydrophobic regions below. Dotted lines represent gaps introduced to maintain maximal alignment of protein sequence.

HTLV or induced by HTLV infection, substantially augment gene expression directed by HTLV LTR sequences. The phenomenon of *trans*-activation distinguishes HTLV from other retroviruses. The unusual structure of the 3′ terminus of HTLV also distinguishes these from most other retroviruses. For this reason, we suggest that the protein encoded by the LOR region may mediate transcriptional changes observed in HTLV-infected cells. In this regard, we note that transcription directed by the HTLV-I LTR is activated to high levels in a cell line, C81-66, that expresses the 42,000-dalton HTLV-I–associated protein but not HTLV *gag*, *pol*, or *env* products (*8*). We further suggest that the HTLV LOR product mediates both the *trans*-activating and transforming effects of HTLV infection. We note that *trans*-acting transcriptional activities have been associated with the transforming genes of other tumor viruses, notably adenovirus and SV40 (*10*, *11*). The existence of a potential transforming function within the HTLV genome may explain the ability of the virus to transform cells in vitro, as well as the absence of specific integration sites in tumor cells and the absence of chronic viremia in target tissues (*12*–*14*). Such a transforming function would differ from that of other retroviruses because, unlike the oncogenes, the sequence that encodes the putative transforming gene will not anneal to the highly conserved cellular sequences (*4*).

Comparison with the bovine leukemia virus genome. We noticed that the 3′ genome of another retrovirus, bovine leukemia virus (BLV), also contains an LOR frame located 3′ to the envelope glycoprotein gene that could encode a protein of a size similar to that of HTLV (*15*, *16*) (Fig. 2). There is evidence for the existence of a subgenomic spliced mRNA species that contains the 3′ open reading frame but not the *gag*, *pol*, and *env* gene sequences in BLV-producing cell lines (*17*).

Although the similarity in structure of the HTLV and BLV proteins is insufficient to indicate that they have a common functional role, the overall similarity in genomic structure, including the location of a 5′ NCR and 3′ LOR frame, and the previously described similarity in protein antigenicity of the two viruses (*1*, *14*) suggests that they are functionally similar. Moreover, there is a similarity in the distribution of hydrophobic and hydrophilic regions of the HTLV and BLV polypeptides. We note that the disease induced by BLV has characteristics similar to those associated with HTLV-I, namely, a long latent period sometimes preceded by persistent lymphocytosis, an absence of chronic viremia in target organs preceding disease, and an absence of preferred integration sites in tumor cells (*18*). These features could be expected of viruses that contain an LOR product mediating transformation.

The biology, structure, and pathology of HTLV and BLV differ from other transforming retroviruses such that we propose that they be considered a new subgroup of retroviruses distinct from both the nonacute transforming viruses that contain only the *gag*, *pol*, and *env* genes and the acute transforming viruses that encode oncogenes.

WILLIAM A. HASELTINE
Dana-Farber Cancer Institute and
Department of Pathology,
Harvard Medical School, and
Department of Cancer Biology,
Harvard School of Public Health,
Boston, Massachusetts 02115
JOSEPH SODROSKI
ROBERTO PATARCA
DEBRA BRIGGS
DENNIS PERKINS
Dana-Farber Cancer Institute
FLOSSIE WONG-STAAL
Laboratory of Tumor Cell Biology,
National Cancer Institute,
Bethesda, Maryland 20205

References and Notes

1. B. J. Poiesz *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 77, 7415 (1980); V. S. Kalyanaraman, M. Sargadharan, P. Bunn, J. Minna, R. C. Gallo, *Nature (London)* 294, 271 (1981); M. Robert-Guroff *et al.*, *J. Exp. Med.* 154, 1957 (1981); M. Yoshida, I. Miyoshi, Y. Hinuma, *Proc. Natl. Acad. Sci. U.S.A.* 79, 2031 (1982).
2. V. S. Kalyanaraman *et al.*, *Science* 218, 571 (1982).
3. M. Popovic, M. G. Sarngadharan, E. Read, R. C. Gallo, *Science* 224, 497 (1984); R. C. Gallo *et al.*, *ibid.*, p. 500; J. Schüpbach *et al.*, *ibid.*, p. 503; M. Sarngadharan, M. Popovic, L. Bruch, J. Schüpbach, R. C. Gallo, *ibid.*, p. 506.
4. M. Seiki, S. Hattori, Y. Hirayama, M. Yoshida, *Proc. Natl. Acad. Sci. U.S.A.* 80, 3618 (1983).
5. G. M. Shaw *et al.*, *ibid.*, in press.
6. G. Franchini, F. Wong-Staal, R. C. Gallo, in preparation.
7. E. P. Gelman, G. Franchini, V. Manzari, F. Wong-Staal, R. C. Gallo, *Proc. Natl. Acad. Sci. U.S.A.* 81, 993 (1984).
8. J. Sodroski *et al.*, *Science* 225, 421 (1984); T. H. Lee *et al.*, in preparation.
9. L. Ratner and F. Wong-Staal, in preparation.
10. J. Brady, J. B. Bolen, M. Radonovich, N. Salzman, G. Khoury, *Proc. Natl. Acad. Sci. U.S.A.* 81, 2040 (1984).
11. N. Jones and T. Shenk, *ibid.* 76, 3665 (1979); J. R. Nevins, *Cell* 26, 213 (1981); A. Berk, F. Lee, T. Harrison, J. Williams, P. A. Sharp, *ibid.* 17, 935 (1979); R. B. Gaynor, D. Hillman, A. Berk, *Proc. Natl. Acad. Sci. U.S.A.* 81 1193 (1984).
12. I. Miyoshi *et al.*, *Nature (London)* 294, 770 (1981); M. Popovic, G. Lange-Wantzin, P. S. Sarin, D. Mann, R. C. Gallo, *Proc. Natl. Acad. Sci. U.S.A.* 80, 5402 (1983); N. Yamamoto, M. Okada, Y. Koyanagi, M. Kannagi, Y. Hinuma, *Science* 217, 737 (1982).
13. B. Hahn *et al.*, *Nature (London)* 305, 340 (1983).
14. R. C. Gallo and F. Wong-Staal, *Blood* 60, 545 (1982).
15. A. Burny, personal communication.
16. N. R. Rice *et al.*, *Virology*, in press.
17. J. Ghysdael, R. Kettmann, A. Burny, *J. Virol.* 29, 1087 (1979).
18. J. F. Ferrer, D. A. Abt, D. M. Bhatt, R. R. Marshak, *Cancer Res.* 34, 893 (1974); P. S. Paul *et al.*, *Am. J. Vet. Res.* 38, 873 (1977); R. Kettmann *et al.*, *J. Virol.* 47, 146 (1983); D. Gregoire *et al.*, *ibid.* 50, 275 (1984); J. Deschamp, R. Kettman, A. Burny, *J. Virol.* 40, 605 (1981); R. C. Gallo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 79, 5680 (1982).
19. A. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* 74, 564 (1977).
20. T. P. Hopp and K. R. Woods, *ibid.* 78, 3824 (1981).
21. We thank A. Burny for providing the sequence of the 3′ region of the BLV, N. R. Rice *et al.* for a preprint of their manuscript, and R. C. Gallo, M. Essex, and J. Coffin for helpful discussions. Supported by an American Cancer Society Directors Grant. J.G.S. was supported by NIH postdoctoral fellowship CA07094.

31 May 1984; accepted 21 June 1984

Sequence of the Envelope Glycoprotein Gene of Type II Human T Lymphotropic Virus

Abstract. *The sequence of the envelope glycoprotein gene of type II human T lymphotropic virus (HTLV) is presented. The predicted amino acid sequence is similar to that of the corresponding protein of HTLV type I, in that the proteins share the same amino acids at 336 of 488 residues, and 68 of the 152 differences are of a conservative nature. The overall structural similarity of these proteins provides an explanation for the antigenic cross-reactivity observed among diverse members of the HTLV retrovirus family by procedures that assay for the viral envelope glycoprotein, for example, membrane immunofluorescence.*

Human T-cell leukemia viruses have been implicated as the etiological agents of several human diseases. The most prevalent type, HTLV-I, is associated with a high incidence of an aggressive form of adult T-cell leukemia (ATLL) and several unusual forms of mycosis fungoides and Sezary syndrome (*1*). A second member of the family, HTLV-II, has been isolated from a patient with benign hairy cell leukemia of T-cell origin (*2*). Recently, a new group of viruses, HTLV-III, was isolated from patients with acquired immune deficiency syndrome (AIDS) (*3*).

The envelope glycoprotein is the major antigen recognized by the serum of persons infected with HTLV (*4*). In this respect HTLV resembles several other retroviruses for which the envelope glycoprotein is typically the most antigenic viral polypeptide (*5*). Moreover, most

/3

Proceedings
OF THE
National Academy of Sciences
OF THE UNITED STATES OF AMERICA

November 1984 | Volume 81, Number 21 | pp. 6565–6908

Table of Contents

Contents

Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 6657–6661, November 1984
Biochemistry

Nucleotide sequence of the 3′ region of an infectious human T-cell leukemia virus type II genome

(retrovirus/DNA sequence/conserved amino acid sequence)

KUNITADA SHIMOTOHNO*, WILLIAM WACHSMAN†, YURI TAKAHASHI*, DAVID W. GOLDE†, MASANAO MIWA*, TAKASHI SUGIMURA*, AND IRVIN S. Y. CHEN†

*National Cancer Center Research Institute, Chuo-ku, Tokyo 104, Japan; and †Division of Hematology and Oncology, Department of Medicine, University of California School of Medicine, Los Angeles, CA 90024

Contributed by Takashi Sugimura, July 12, 1984

ABSTRACT **The nucleic acid sequence of the 3′ region of human T-cell leukemia virus type II (HTLV-II) proviral DNA was determined using a HTLV-II proviral clone that could be recovered as infectious, transforming virus. The sequence data indicate a region of unknown function of ≈1.6 kilobase pairs in the 3′ region, analogous to the X region previously identified in human T-cell leukemia virus type I (HTLV-I). Three overlapping open reading frames are present in the X region of HTLV-II. One of these open reading frames, Xc, is most likely to encode a protein product, because it has greater predicted amino acid sequence homology (78%) with the X-IV region of HTLV-I and a greater percentage of its base differences with X-IV at the third nucleotide position of codons than do the other open reading frames. Sequences of the X-region that include the open reading frames are conserved in two deletion mutants of HTLV-II, which are associated with a subline of Mo cells with a decreased dependence on fetal bovine serum.**

Human T-cell leukemia viruses (HTLV) are associated with certain forms of human leukemias and lymphomas (1–5). At least two types of HTLV have been identified. HTLV type I (HTLV-I) is endemic to various regions of the world and is often associated with aggressive leukemias/lymphomas of mature T lymphocytes (3–5). HTLV type II (HTLV-II) was found in a single patient (Mo) with a T-cell variant of hairy-cell leukemia (1, 2, 6). This patient is alive and well 8 yr after splenectomy.

Both HTLV-I and HTLV-II transform normal human peripheral blood or cord blood T lymphocytes *in vitro* (7–10). These virus-transformed T cells have a helper-inducer phenotype similar to that of leukemic cells in patients with HTLV-associated disease. Elucidation of the mechanism of *in vitro* transformation is relevant to the process of leukemogenesis. However, the regions of the HTLV genome necessary for transformation have not been identified. Nucleic acid sequence analysis of the complete HTLV-I genome revealed a region at the 3′ locus of the genome with no known function and without precedent in animal retroviruses other than bovine leukemia virus (11). This region, referred to as X, is suspected to encode protein(s) involved in the process of transformation. The X region does not cross-hybridize with normal human cellular DNA sequences and, therefore, does not encode a retroviral oncogene.

HTLV-II has *in vitro* biological properties similar to but only limited homology with HTLV-I as determined by hybridization of the genomes and nucleic acid sequencing of the long terminal repeat (LTR) (12, 13). By nucleic acid sequence analysis we have identified a region comparable to X in an infectious and transformation-competent molecular clone of HTLV-II. The homology between the two viruses in this region was determined.

MATERIALS AND METHODS

Sequencing of HTLV-II DNA. Bacterial plasmid pH6-B3.5, which contains *env*, X, and a part of the LTR of HTLV-II, was used as a source of DNA. pH6-B3.5 was subcloned from a cloned infectious HTLV-II provirus, λH6. The sequencing method of Maxam and Gilbert was applied to 5′- or 3′-end-labeled DNA fragments obtained by digestion of the DNA with restriction enzymes (14). Both strands of the DNA were sequenced.

Comparison of Nucleotide and Amino Acid Sequences. Nucleotide or amino acid sequence homology was assessed using a computer program developed by Japan Soft Development.

Transfection of HTLV-II Proviral DNA. The procedures used for transfection of lymphoid cells were described previously (15).

Materials. Restriction enzymes were purchased from Takara Shuzo (Kyoto, Japan), Bethesda Research Laboratories, or New England Biolabs. Polynucleotide 5′-hydroxyl-kinase and the large fragment of DNA polymerase I were from Boehringer Mannheim and Takara Shuzo, respectively. Radiolabeled nucleotides were from Amersham.

RESULTS

Nucleic Acid Sequence Analysis of the 3′ Region of the HTLV-II Genome. Previous nucleic acid sequence analysis of HTLV-I revealed four potential open reading frames beginning with methionine codons in the 3′ region of *env* (11). However, as the sequenced provirus was not recovered as an infectious virus it may not represent the genome of a replication- and transformation-competent HTLV-I. Therefore, we first determined whether an apparently complete HTLV-II provirus clone, λH6, could be recovered as infectious virus capable of transforming normal human T lymphocytes.

Since HTLV-II can replicate in some B-lymphoblastoid cell lines, we used a B-cell line for HTLV-II DNA transfection. The HTLV-II provirus was subcloned into the plasmid vector pSV2-neo. Protoplasts of *E. coli* HB101 containing the HTLV-II subclone, pH6-neo (Fig. 1*A*), were fused with WIL-2 cells and antibiotic G418-resistant clones of cells were subsequently selected. Of these G418-resistant B-cell clones, ≈25% expressed viral p19 and p24 antigens, as determined by indirect immunofluorescence (Fig. 1*B*), and viral RNA which was correctly initiated from the cap site of the LTR (15). These B cells were lethally irradiated and cocultivated with normal human peripheral blood lymphocytes as

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HTLV-I and HTLV-II, human T-cell leukemia virus types I and II; LTR, long terminal repeat; kbp, kilobase pair(s).

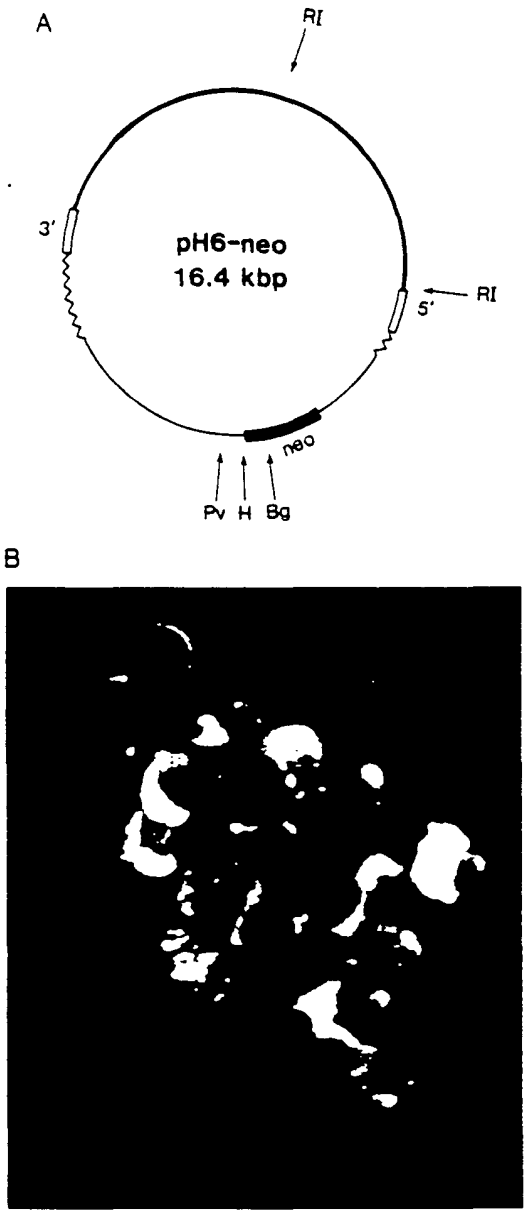


FIG. 1. Transfection and expression of HTLV-II proviral DNA in human B cells. (*A*) The plasmid pH6-neo used for transfection is shown schematically. The complete HTLV-II provirus of λH6 (12) between flanking *Hin*dIII sites in cellular DNA was subcloned into the *Eco*RI site of the plasmid vector pSV2-neo. The DNAs cleaved with *Hin*dIII or *Eco*RI were treated with nuclease S1 to blunt the ends and were ligated. WIL-2 cells were transfected with pH6-neo by spheroplast fusion (15) and stable transformants were propagated. The thick line and open boxes represent the provirus. Zig-zag lines show cellular flanking sequences. The thin lines and shaded box represent pSV-neo sequences. *Bgl* II (Bg), *Hin*dIII (H), *Eco*RI (RI), and *Pvu* II (Pv) sites are indicated. (*B*) One of the stable B-cell transformants that express HTLV-II p24 antigens is shown. Fixed cells were treated with rabbit antisera directed against viral p24 antigens and visualized by indirect immunofluorescence.

described to test for HTLV-II transformation (10). Cell lines with the T-helper surface-antigen phenotype were established and shown to be infected with HTLV-II. No B-cell markers (surface membrane immunoglobulin, Epstein–Barr-virus nuclear or capsid antigen) were detected in the transformed peripheral blood cells. The nucleic acid sequence of the λH6 provirus therefore represents the genome of a replication- and transformation-competent HTLV-II.

The nucleotide sequence of the 3' region of HTLV-II is shown in Fig. 2; *env* ends one base before the first nucleotide of Fig. 2 and the X region of HTLV-II extends to position 1559.

Sequence Homology Between the X Regions of HTLV-II and HTLV-I. There is considerable nucleotide sequence homology between the X regions of HTLV-II and HTLV-I. Most of this sequence homology is in the 3' two-thirds of the X region that is coincident with the position of the major open reading frames in the X region (see below). In this region there is 75% nucleic acid sequence homology. In contrast there is only 33% sequence homology in the 5' one-third of the X region. The 5' region corresponds to the X-I region of HTLV-I (11). No long open reading frames are present in this region of HTLV-II.

Three open reading frames with overlapping sequences were identified in the nucleotide sequence of HTLV-II. These open reading frames Xa, Xb, and Xc correspond to the open reading frames X-II, X-III, and X-IV of HTLV-I (11). The derived amino acid sequence homologies for the codons between flanking termination codons are 62, 61, and 78% for Xa/X-II, Xb/X-III, and Xc/X-IV, respectively. Xc and X-IV share a stretch of 335 corresponding amino acid codons uninterrupted by termination codons (Fig. 3), compared with 96 for Xa/X-II and 145 for Xb/X-III. Furthermore, there is 82% amino acid homology in the 112 codons upstream of the first conserved methionine codon of Xc/X-IV, indicating that the Xc region may encode a fused protein whose initiation codon is in a different region of the genome.

Frequency of Base Changes at the Third Nucleotide Position in the Open Reading Frames. The frequency of base changes between divergent sequences in each of the three positions in amino acid codons has been used as a measure of evolutionary conservation and, therefore, functional significance of an open reading frame: the greater the frequency of third nucleotide changes relative to the first and second, the more likely that the reading frame encodes a protein that has been conserved during evolution.

The frequency of base changes in the open reading frames of HTLV was calculated by comparing the sequences of the corresponding open reading frames in HTLV-I and HTLV-II. The number of mismatched bases between the Xa region (from nucleotides 530 to 817) in HTLV-II and the corresponding region in the open reading frame for X-II in HTLV-I is 58 mismatched nucleotides out of 288 nucleotides (only amino acid codons between termination codons are included in this calculation). Fourteen percent of these mismatches are at the third position of codons in this reading frame. In the Xb/X-III region, 28% of the mismatched bases are at the third position. However, for Xc/X-IV, 66% of the mismatched bases are located at the third position, a significantly greater frequency than that of either of the other two reading frames. Thus, Xc is most likely to encode a protein product.

Conservation of the 3' Region of HTLV-II DNA in Deletion Mutants. Molecular cloning and characterization of HTLV-II DNA from Mo cells demonstrated the presence of three forms of HTLV-II proviruses in these cells. The largest cloned provirus represents the complete replication-competent genome of HTLV-II as evidenced by the recovery of infectious, transforming HTLV-II by DNA transfection. The other two forms of HTLV-II DNA were defective, having large internal deletions of the viral genome. However, their LTRs were intact and both defective genomes could be packaged as infectious virus (12). The defective viruses are associated only with a subline of the HTLV-II-infected Mo cells that has growth properties distinct from the original Mo cell line: these cells have a decreased dependence on fetal bovine serum and clone spontaneously in methylcellulose and by limiting dilution.

Restriction enzyme analysis showed that the larger defec-

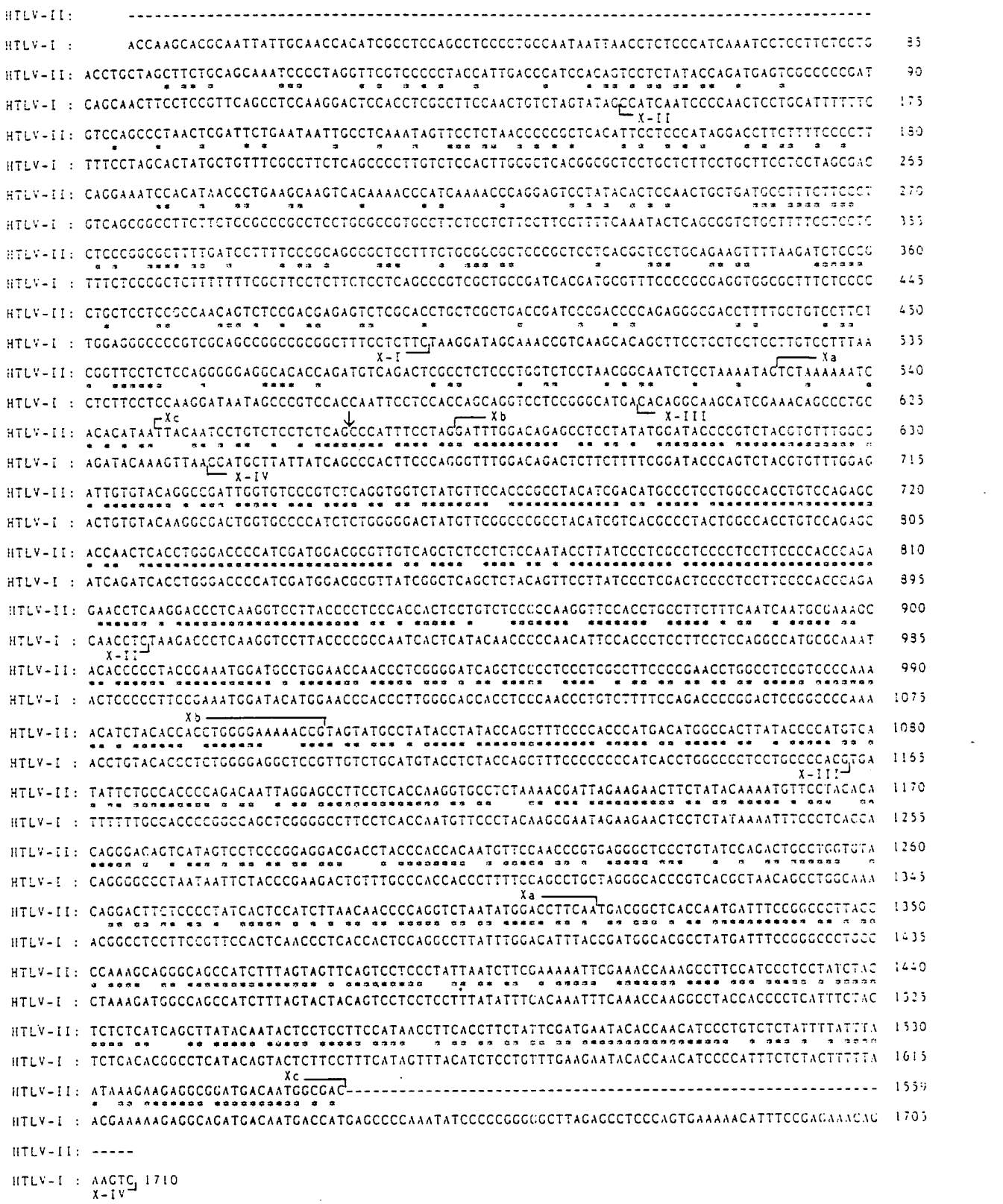


FIG. 2. Nucleotide sequence homology between the X regions of HTLV-II and HTLV-I proviruses. The nucleotide next to the 3′ end of the *env* gene is designated nucleotide 1. The open reading frames Xa, Xb, and Xc in HTLV-II and X-I, X-II, X-III, and X-IV in HTLV-I are shown. The 5′ portion of the X-I open reading frame is in *env*. Xc and X-IV end in the LTRs. A putative splice acceptor site is indicated by an arrow. Asterisks indicate nucleotides that are identical in the two sequences.

tive genome, typified by clone H9 (Fig. 4), has a deletion of ≈2.0 kilobase pairs (kbp) with conservation of ≈5.0 kbp in the 5′ region and ≈2.0 kbp in the 3′ region. The smaller defective provirus, typified by clone H2, has a deletion of nearly the entire internal sequence of HTLV-II. Excluding the LTR, <2.0 kbp of the sequence is conserved. Detailed restriction enzyme analysis demonstrated that most of the conserved sequence in the 3′ region is the X region of HTLV-II.

```
HTLV-II : LQSCLLSAHFLG FGQSLLYGYP VYVFGDCVQA DWCPVSGGLC STRLHRHALL ATCPEHQLTW   62
          ******* * ****** *** ********* **** ***** * ******** ******* **
HTLV-I  : --PCLLSAHFPG FGQSLLFGYP VYVFGDCVQG DWCPISGGLC SARLHRHALL ATCPEHQITW   60

HTLV-II :   DPIDGRVVSS PLQYLIPRLP SFPTQRTSRT LKVLTPPTTP VSPKVPPAFF QSMRKHTPYR  122
            ******* *  ** ****** ******** * ******* *    *  ** *  * *** * *
HTLV-I  :   DPIDGRVIGS ALQFLIPRLP SFPTQRTSKT LKVLTPPITH TTPNIPPSFL QAMRKYSPFR  120

HTLV-II :   NGCLEPTLGD QLPSLAFPEP GLRPQNIYTT WGKTVVCLYL YQLSPPMTWP LIPHVIFCHP  182
            **  *****   ** * ** * ****** **  **  *** ** ****** *** * *********
HTLV-I  :   NGYMEPTLGQ HLPTLSFPDP GLRPQNLYTL WGGSVVCMYL YQLSPPITWP LLPHVIFCHP  180

HTLV-II :   RQLGAFLTKV PLKRLEELLY KMFLHTGTVI VLPEDDLPTT MFQPVRAPCI QTAWCTGLLP  242
            ******* * * ** ***** * * ** * **** **** *** ***   *** ****
HTLV-I  :   GQLGAFLTNV PYKRIEELLY KISLTTGALI ILPEDCLPTT LFQPARAPVT LTAWQNGLLP  240

HTLV-II :   YHSILTTPGL IWTFNDGSPM ISGPYPKAGQ PSLVVQSSLL IFEKFETKAF HPSYLLSHQL  302
             ** ****** **** ** ** **** ** ** **** ***   ** ** ***  *** **** *
HTLV-I  :   FHSTLTTPGL IWTFTDGTPM ISGPCPKDGQ PSLVLQSSSF IFHKFQTKAY HPSFLLSHGL  300

HTLV-II :   IQYSSFHNLH LLFDEYTNIP VSILFNKEEA DDNGD                            337
            ******* ** *** ******  * ***  ** ***
HTLV-I  :   IQYSSFHSLH LLFEEYTNIP ISLLFNEKEA DDNDHEPQIS PGGLEPPSEK HFRETEV     357
```

FIG. 3. Homology of predicted amino acid sequences encoded by the open reading frames Xc (HTLV-II) and X-IV (HTLV-I). Asterisk indicate identical amino acids in the two sequences. Amino acids are represented by standard one-letter abbreviations (16).

Furthermore, the deletion endpoints occur at the 5′ end of the X region, upstream of the large open reading frames (Fig. 4).

DISCUSSION

We have sequenced a 3′ region of the HTLV-II genome of ≈1.6 kbp with an unknown function. HTLV-II resembles HTLV-I in a number of its properties, including biological functions, such as lymphoid target-cell specificity and T-cell transformation (1, 2, 6–10), and conservation of important structural features within the LTR (13). The X region represents another common structural feature that is present in the genomes of both HTLV types. The X region of HTLV-II has 61% sequence homology with the X region of HTLV-I, and shows homology as great as 75% in the region encompassing the 3′ two-thirds of the X region. The sequence conservation in this part of the X region in both types of HTLV strongly suggests that the X region serves an important function in virus replication and/or transformation. Three large open reading frames with overlapping sequences are present in the HTLV-II X region. If they began with initiation codons, these open reading frames would be sufficient to encode proteins of 15,100, 23,700, and 24,500 daltons. Although the predicted amino acid sequence of HTLV-II in these regions shows ≈60% homology with those of the X-II and X-III regions and 75% homology with that of the X-IV region of HTLV-I, the positions of initiation and termination codons would result in proteins of different predicted sizes.

Comparison of the LTR sequences of HTLV-II and HTLV-I indicates that these two viruses are only distantly related. It is likely that the two viruses evolved from a single ancestral virus and have retained common sequences that are important for replication (13). Of the corresponding open reading frames in the two viral genomes, Xc and X-IV share the greatest sequence homology, the longest stretch of contiguous codons uninterrupted by termination codons, and the greatest frequency of third-position differences relative to first- and second-position differences. Therefore, it is likely that Xc in HTLV-II and X-IV in HTLV-I encode functional proteins.

The high predicted amino acid homology and relatively high frequency of third-position differences holds true for 112 codons in Xc/X-IV located upstream of the first methionine codon. Therefore, it is probable that Xc is translated as a fused protein from a spliced mRNA. In this regard, it is interesting to note that a potential splice acceptor site is lo-

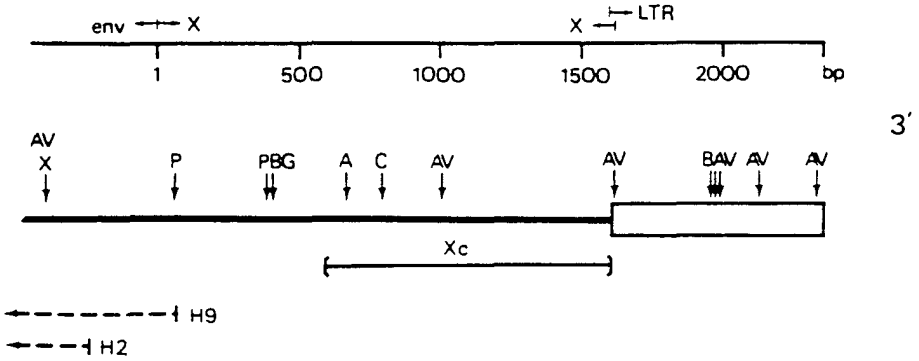


FIG. 4. Location of deletion endpoints in the HTLV-II H-2 and H-9 clones. The restriction enzyme map of the HTLV-II X region and 3′ LTR (represented by a box) is illustrated. Restriction enzyme sites are shown for *Acc* I (A), *Ava* I (AV), *Bam*HI (B), *Bgl* II (BG), *Cla* I (C), *Pst* I (P), and *Xho* I (X). Boundaries between *env* and X and between X and the LTR are indicated. The 3′ deletion endpoints of the defective H-2 and H-9 clones, as determined by restriction enzyme mapping and subsequent hybridization analysis (unpublished data) are shown in reference to the λH-6 restriction enzyme map. The location of Xc is denoted by a bracketed line. bp, Base pairs.

cated near the 5′ end of the Xc/X-IV region at nucleotide position 570.

Since animal retroviruses for which nucleic acid sequence information is available do not have sequences comparable to the X region of HTLV, it is likely that this X region has a unique function in viral replication and cellular transformation. The retention of the open reading frames in the X region of the deletion mutants of HTLV-II may be relevant to that region's potential function in transformation, particularly since the deletion mutants are present only in a subline of the Mo cells having much less stringent growth requirements than the parental Mo cells. Identification of the proteins encoded by the X region and X-region-specific mRNAs will be necessary to determine the significance of the X region.

Note Added in Proof. While this work was in press, Haseltine *et al.* (17) published a sequence of the 3′ region of HTLV-II that differs from that presented here at six nucleotide positions in the Xc region; four of these differences result in amino acid changes. These differences from our data may be due to sequence differences in the two provirus clones used for analysis; the significance, if any, of these differences must await demonstration of the infectivity of the cloned HTLV-II provirus used for sequencing (18).

We thank J. McLaughlin and S. Quan for technical assistance, J. Gasson and T. Sekiya for comments on the manuscript, and T. Gojobori for helpful discussions. This work was partly supported by Grants-in-Aid for Cancer Research from the Ministry of Education, Science, and Culture and from the Ministry of Health and Welfare of Japan; U.S. Public Health Service Grants CA 30388, CA 32737, CA 09297, and CA 16042 awarded by the National Cancer Institute; and Grant PF-2182 from the American Cancer Society.

1. Saxon, A., Stevens, R. H. & Golde, D. W. (1978) *Ann. Intern. Med.* **88**, 323–326.
2. Kalyanaraman, V. S., Sarngadharan, M. G., Robert-Guroff, M., Miyoshi, I., Blayney, D., Golde, D. W & Gallo, R. C. (1982) *Science* **218**, 571–573.
3. Poiesz, B. J., Ruscetti, F. W., Gazdar, A. F., Bunn, P. A., Minna, J. D. & Gallo, R. C. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 7415–7419.
4. Hinuma, Y., Nagata, K., Hanaoka, M., Nakai, M., Matsumoto, T., Kinoshita, K.-I., Shirakawa, S. & Miyoshi, I. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6476–6480.
5. Blattner, W. A., Kalyanaraman, V. S., Robert-Guroff, M., Lister, T. A., Galton, D. A. G., Sarin, P. S., Crawford, M. H., Catovsky, D., Greaves, M. & Gallo, R. C. (1982) *Int. J. Cancer* **30**, 257–264.
6. Saxon, A., Stevens, R. H., Quan, S. G. & Golde, D. W. (1978) *J. Immunol.* **120**, 777–782.
7. Miyoshi, I., Kubonishi, I., Yoshimoto, S., Akagi, T., Ohtsuki, Y., Shiraishi, Y., Nagata, K. & Hinuma, Y. (1981) *Nature (London)* **294**, 770–771.
8. Yamamoto, N., Okada, M., Koyanagi, M. & Hinuma, Y. (1982) *Science* **217**, 737–739.
9. Popovic, M., Sarin, P. S., Robert-Guroff, M., Kalyanaraman, V. S., Mann, D., Minowada, J. & Gallo, R. C. (1983) *Science* **219**, 856–859.
10. Chen, I. S. Y., Quan, S. G. & Golde, D. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 7006–7009.
11. Seiki, M., Hattori, S., Hirayama, Y. & Yoshida, M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3618–3622.
12. Chen, I. S. Y., McLaughlin, J., Gasson, J. C., Clark, S. C. & Golde, D. W. (1983) *Nature (London)* **305**, 502–505.
13. Shimotohno, K., Golde, D. W., Miwa, M., Sugimura, T. & Chen, I. S. Y. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1079–1083.
14. Maxam, A. M. & Gilbert, W. (1980) *Methods. Enzymol.* **65**, 499–560.
15. Chen, I. S. Y., McLaughlin, J. & Golde, D. W. (1984) *Nature (London)* **309**, 276–278.
16. IUPAC-IUB Commission on Biochemical Nomenclature (1968) *Eur. J. Biochem.* **5**, 151–153.
17. Haseltine, W. A., Sodroski, J., Patarca, R., Briggs, D., Perkins, D. & Wong-Staal, F. (1984) *Science* **225**, 419–421.
18. Gelmann, E. P., Franchini, G., Manzari, V., Wong-Staal, F. & Gallo, R. C. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 993–997.